

Expectancy-value based achievement motivations and their role in student learning

Dirk T. Tempelaar

© Dirk T. Tempelaar, Maastricht, 2007
ISBN: 978-90-5278-625-4
Omslagontwerp: Bas van Kesteren
Druk en bindwerk: Datawyse bv, Maastricht

Expectancy-value based achievement motivations and their role in student learning

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht
op gezag van de rector magnificus
prof. mr. G. P. M. F. Mols
volgens het besluit van
het College van Decanen
in het openbaar te verdedigen

op donderdag 31 mei 2007
om 16.00 uur

door
Dirk Tiemen Tempelaar



Promotor

Prof. Dr. W. H. Gijselaers;

Copromotor

Dr. S. Schim van der Loeff;

Beoordelingscommissie

Prof. Dr. F. Palm (voorzitter);

Prof. Dr. M. Berger;

Prof. Dr. M. S. R. Segers, Universiteit Leiden;

Prof. Dr. J. Vermunt, Universiteit van Utrecht;

Prof. Dr. C. van der Vleuten.

Table of contents

Chapter 1: Presage	p. 1
Chapter 2: Puzzles in statistical reasoning Based on: Tempelaar, D. T., Gijsselaers, W. H., & Schim van der Loeff, S. (2006). Puzzles in statistical reasoning. <i>Journal of Statistics Education</i> , 14(1).	p. 13
Chapter 3: Attitudes toward statistics and prior reasoning abilities Based on: Tempelaar, D. T., Schim van der Loeff, S., & Gijsselaers, W. H. (2007). A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities, and course performance. Manuscript under revision for publication in <i>Statistics Education Research Journal</i> .	p. 45
Chapter 4: Commonalities in attitudes and beliefs toward different academic subjects Based on: Tempelaar, D. T., & Nijhuis, J. F. H. (2007). Commonalities in attitudes and beliefs toward different academic subjects. In M. K. McCuddy, H. van den Bosch, J. W. B. Martz, A. V. Matveev, & K. O. Morse (Eds.), <i>Educational Innovation in Economics and Business X: The challenges of educating people to lead in a challenging world</i> (pp. 225-250). Berlin: Springer.	p. 81

Chapter 5: Student achievement motivations in business subjects **p. 109**

Based on: Tempelaar, D. T., Gijssels, W. H., Schim van der Loeff, S., & Nijhuis, J. F. H. (2007). A structural equation model analyzing the relationship of student achievement motivations and personality factors in a range of academic subject-matter areas. *Contemporary Educational Psychology*, 32(1), 105-131.

Chapter 6: Variations in student achievement motivations **p. 151**

Based on an abridged version of: Tempelaar, D. T., Schim van der Loeff, S., & Gijssels, W. H. (2007). Variations in achievement motivations over business subjects. Manuscript submitted to *Academy of Management Learning & Education*.

Chapter 7: Postscript **p. 177**

Samenvatting **p. 183**

Over de auteur **p. 189**

Dankwoord

Ooit was het 't beoogde onderwerp van mijn proefschriftonderzoek: toepassingen van dynamische factoranalyse. Dat was begin jaren '80, Groningen, met de primeur van een eerste ZWO-subsidie voor gamma-onderzoek op zak, temidden van een groep aanstekelijk enthousiaste mathematische systeem-theoretici: Christiaan Heij, Hans Nieuwenhuis, Pieter Otter, Jan Willems. Van jongs af aan opgegroeid in omgevingen waarin debat en betoog belangrijke communicatievormen zijn en de leefwereld veel wetenschappelijke trekjes vertoont, was dit zeker de droomstart van een begeerde academische carrière. Een start die overigens op zijn beurt weer in het niet viel bij de uitdaging die zich al kort daarna aandiende: het meebouwen aan een nieuwe opleiding in Maastricht. Het aannemen van die uitdaging begin '84 impliceerde wel een drastische wijziging van mijn werkzaamheden: het ene onderzoeksproject werd vervangen door een veelheid van andere projecten, die naast en na elkaar zo concurreerden om aandacht, dat onderzoek er voorlopig even niet tussen paste. Alhoewel ontegenzeggelijk niet slim vanuit een loopbaanperspectief, ben ik de vakgroep, c.q. Franz Palm als werkbaas & semi-permanente vakgroepvoorzitter, dankbaar dat de ruimte bestond om met zo veel verschillende, belangrijke maar vooral ook leuke klussen bezig te zijn. Maar weinigen stralen in eigen gedrag, en in houding naar anderen, zo expliciet uit dat het belang van faculteit en universiteit altijd bovengeschied is aan dat van vakgroep of persoon. Het aanlokkelijke van al die projecten kwam natuurlijk niet alleen voort uit de uitdagende taakstelling, maar in belangrijke mate ook uit de zeer plezierige samenwerking met nagenoeg al mijn collega's die ik in de verschillende projecten heb ontmoet. Met het risico velen tekort te doen, zou ik graag een expliciet dankwoord richten aan de UM collega's waarmee ik in de volgende, ongeveer chronologisch opgesomde 'projecten', heb mogen samenwerken:

- Bestuur, in de vorm van één periode FB (als vertegenwoordiger van de jonge honden), twee perioden UR en een ongeteld aantal perioden FR: René Verspeek, Wilma Coenegrachts, Wil Albeda, Franz Palm, Bram Beek, Erik de Regt, Louis Berkvens, Arjan Blokland.
- FdEWB curriculumopbouw, met als onbetwiste climax de 'lijstencommissie' en alle beraadslagingen in de Commissie Economie en daarbuiten, over doorlopende rode draden in het

verplichte onderwijs: Piet Keizer, Paul van Loon, Tom van Veen, Geert Woltjer, Joan Muysken, Hans Kasper, Jos Lemmink.

- Project 'Macs voor studenten/Océ netwerk/ACO/Computerplatform': René Verspeek, Joan Muysken en Paul Hick waren daar mijn eerste sparringpartners voor al te ambitieuze plannen, later vervolgd door Jean-Paul Beusen, Hub van Kan, Bahram Shahbaz Moradi, Geert Woltjer, Mark Arts, Laury Bollen en anderen;
- Project UniversiteitsBibliotheek met in de loop der tijd een tiental commissies en één heus hoogtepunt, het m.i. meest geslaagde UM nieuw/verbouwproject, te weten de bibliotheek binnenstad: John Gilbert, Henk van Hoogen, Ron Aardening, Charles Bollen, Nico Cobben, Marianne de Ruwe, Paul Kunst, Tom van Veen, Erik de Regt, Wim Swaan en Rita Walczuch.
- Het 'toetsproject', zijnde de opeenvolging van VoortGangsToetsing, het Boek Toetsconstructie, en de OverAllToetsing: Winand Wijnen, Mien Segers, Tom van Veen, Jan Nijhuis.
- Project 'EDINEB' en de daaruit voortvloeiende Peruaanse, Litouwse en Russische avonturen: Wim Gijselaers, Ellen Nelissen, Jeannette Hommes en Piet Keizer.
- Het allesoverheersende project 'QM', inclusief dik 20 jaar experimenteren met 'high tech' onderwijsinnovaties en alle discussies daarover: Denis de Crombrughe, Sybrand Schim van der Loeff, Christian Kerckhoffs als onvermijdelijke 'tegenvoeter', Stan van Hoesel, Hans Peters, Lutz Krebs, Henk-Jan van der Kamp, Tjaart Imbos, Nick Broers, Luc Budé, en sinds enige jaren mijn UCM collega's.
- Het uit de kluiten gegroeide DSS-project, met steeds onverwachte steun uit besliskundige hoek: Jan Willem Goossens, Anton v.d. Kraaij, Alexander Grigoriev en Joyce van Loon.
- Elektronisch leren, in de vele gedaanten van Eleum, WebSpijkeren 1 en 2, Emerge en NKBW waarin deze ontwikkeling zich heeft opgewerkt tot de grote missie van zeker Bart, Mark en mezelf: Bart Rienties, Mark Arts, Jean-Paul Beusen, Hans de Graaff, Maarten van Wesel en Peter Verheijen.
- Universitair hockey in de bedrijfscompetitie (naar mijn weten het enige sportieve visitekaartje van de Universiteit in een reguliere sportcompetitie): Sybrand Schim van der Loeff, René Verster, Piet Eichholtz, Ralph Schoneman, Aalt Willem Heringa, Fred Stevens,

Gerard Swaan, Jean Marie Willems, Dick Veul, en lang geleden, Piet Keizer.

- Het project waaraan ik, als rechtgeaard KE-er, zeker meer heb geprofiteerd, dan heb bijgedragen: het 'collegialiteitsproject'. Dank aan Yolanda Paulissen, Karin van den Boorn, Haydeé Hallmanns, en, op hun geheel eigen manier, aan Eugene Schenk en Gerard Pfann.
- Het project 'niet-universitair pgo', ook wel geheten Maastrichts Montessori-onderwijs, in de gestalten van de basisscholen Montessori Binnenstad en Montessori West / Rotonde, onze stichting SVM, en het Montessori College / Vorming: Huub Stassen, Dré Gybels, Ben Boden, Math Korpershoek, Fred Kelpin, Jos v. d. Wiel, Paula Deumens, Aalt Willem Heringa, Ineke Wolfhagen, Winand Wijnen.

En dan het laatste project: de terugkeer naar het veld van het onderzoek. Met een onderwerp dat verrassend veel gelijkenis vertoont met dat van 25 jaar eerder: toepassingen van factoranalyse. Alleen nu niet meer dynamisch en, veel belangrijker, binnen een ander toepassingsgebied: de econometrie is verruild voor de psychometrie, of nog preciezer, modellen van leerprocessen uit de onderwijspsychologie. Door die onderwerps-verandering was het project van te voren onfeilbaar geworden: wanneer de hobby van het kwantitatief modelleren gecombineerd kan worden met de hobby van het onderwijs, moet succes wel gegarandeerd zijn. Desondanks was het geen eenvoudig project, waarin de twee promotores Wim en Sybrand heel verschillende maar zeer cruciale rollen hebben gespeeld: dan weer eens eendrachtig opponerend tegen zoveel ongefundeerde eigenzinnigheid, dan weer van de eigen flanken opererend in een poging mathematische precisie te versterken, de zo graag gebezigde schrijfstijl van spanningsboogopbouw de kop in te drukken, of gewoon meer oog te hebben voor de kenmerken van lezers en recensenten van wetenschappelijke tijdschriften. Jullie rol en de vasthoudendheid waarmee die tot letterlijk de laatste dag is volgehouden, wordt ten eerste geapprecieerd. Dat laatste geldt ook de rol van mijn twee vaste sparringpartners van afgelopen jaren: Bart en Jan.

Een latere promotie heeft ook wel z'n voordelen. Eén ervan is dat een ander toch vaak omvangrijk project, het project gezin, er niet bovenmatig onder heeft hoeven lijden. Wanda, Jernst en Ynske hebben die kritische

coach langs de lijnen van het hockeyveld weinig zaterdagen hoeven missen; wellicht hadden ze het anders gewenst. De kosten zijn nu geheel afgewenteld op slechts één iemand: Concha. Met veel dank wordt het geduld, de steun én het tegenwicht aanvaard; voor mij dé illustratie van het cruciale belang van diversiteit in elk leerproces.

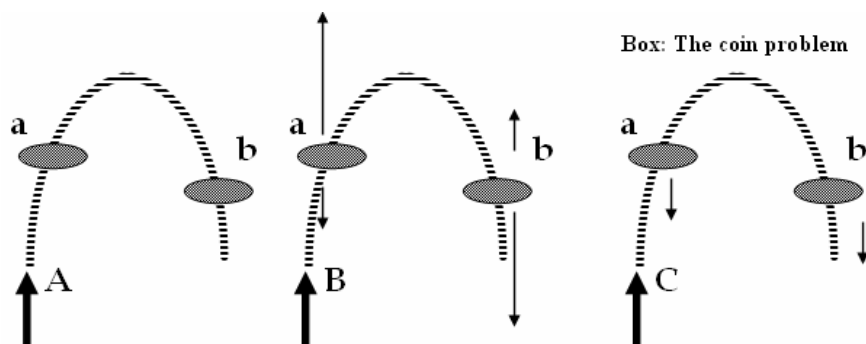
Dirk Tempelaar
Maastricht, 2007

Chapter 1

PRESAGE

1. INTRODUCTION: LEARNING BASED ON PRIOR KNOWLEDGE

Prior knowledge is generally regarded as the most influential single determinant of academic learning processes, and as best predictor of any assessment of the outcomes of such a learning process (Alexander, 2006; Bransford et al., 2000). Educational research that grounds this finding has typically focussed on the most intuitive interpretation of prior knowledge: academic knowledge in the same domain, such as partial but correct knowledge. However, this is a limited interpretation of prior knowledge, since at least in some domains other types of prior knowledge, that we shall refer to as naïve theories or misconceptions, may exert their influence on the learning process and its outcomes. The coin problem (Bruer, 1997; Clement, 1982) described in the following Box provides an illustration of this other kind of prior knowledge.



A coin is tossed in the air and follows the path shown in panel A. In which panel, B or C, do the arrows correctly show the forces acting on the coin on the upward (a) and downward (b) parts of its trajectory?
Source: Bruer (1997, p. 127)

Most readers will immediately recognize in this problem the assessment of Newton's First and Second Laws of Motion. The First Law says that a body moves uniformly in a straight line if no force acts on that body, and the Second Law states that if such a force acts on a body, the force equals the product of body mass and acceleration. If the problem were to be incorporated in a mechanics exam for students in either secondary or higher education, most students can be expected to give the appropriate answer to this problem. But if this problem is posed to students, that have studied mechanics, outside any testing context, surprisingly the vast majority of students will answer in the following way, labelled by Clement (1982, p. 67) as the 'typical incorrect answer': *"While the coin is on the way up, the 'force from your hand' pushes up on the coin. On the way up it must be greater than gravity, otherwise the coin would be moving down."* This typical incorrect answer, that corresponds to panel B, is remarkably dominant. Clement reports that even engineering students, after having successfully completed a physics course in mechanics for scientists and engineers, in large majority (70%) give incorrect answers. In doing so, they fall back to what Bruer calls naïve theories or misconceptions: informal, self-acquired science knowledge, inconsistent with formal science. These proto-scientific concepts and schemas, developed in childhood, have severe consequences for science instruction. In addition to, or rather, before learning formal theories, it should aim at the unlearning of naïve theories. The required efforts for the latter part will in general surpass those needed for the first part.

Is physics unique in the existence of naïve theories? No, some other domains do have their share, and it is not that difficult to come up with some natural candidates of domains facing the risk of the existence of naïve theories. Long time lags between natural learning, taking place outside school, and formal school learning, creates ample opportunities for naïve theories to develop. Mechanics, exemplified by falling objects like in the coin problem, is an obvious demonstration of such a time lag: confronted with gravitation in early childhood, pupils have to wait till middle or high school to learn Newton's first and second law. Statistics and probability theory, including reasoning in situations under uncertainty, constitutes another example that in terms of the length of the time lag even surpasses mechanics: where data sets and probabilities enter the stage again in early childhood ('we have not eaten pancakes for days; good chance we will have them today'), the confrontation with formal theories is mostly not earlier than in university. Naïve statistical theories that develop during this long episode do not only impact the learning of statistics itself, but also the learning in other domains in which statistical reasoning is applied to explain the behaviour of people. Finance is an exponent of such a domain: standard finance theory is built on several principles (arbitrage principle, portfolio principle, capital asset pricing theory and option-pricing theory; see e.g. Statman, 1999) that all assume rational behaviour of investors. However, there is a wealth of evidence that investors act irrationally in situations under risk (Bernstein, 1996) and this irrationality even follows predictable patterns, described by the theory of behavioural finance (Statman). Behavioural finance is based on frequently observed naïve statistical theories, and its existence in itself makes clear that the development of naïve statistical theories is an important object of research. This thesis will focus on the explanation of the development of such naïve statistical theories.

Research into naïve statistical theories is best known under the heading statistical reasoning (Garfield & Ahlgren, 1988; Shaughnessy, 1992). A milestone in the development of this cognitive psychological research is the work by Tversky and Kahneman, collected in edited volumes as Kahneman, Slovic, and Tversky (1982), Kahneman and Tversky (2000), Tversky and Shafir (2004). This research classifies different types of common fallacies in judgements under uncertainty. The conjunction fallacy, described in the next Box, is a typical example of such a type.

Box: The conjunction fallacy

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstration.

Please rank the following statement by their probability, using 1 for the most probable and 3 for the least probable.

- Linda is active in the feminist movement.
- Linda is a bank teller.
- Linda is a bank teller and is active in the feminist movement.

Source: Tversky and Kahneman (1982, p. 92)

Remark: in its original format, the problem contains eight statements, with another five serving the role of distractors.

The title of the fallacy more or less betrays its essence: statement c is the conjunction of statements a and b, and therefore, the probability of c is in general smaller than any of the individual probabilities of statements a and b. The statements of targets, as Tversky and Kahneman (1982) call them, are constructed in a specific way: one is representative of the provided description of the person (a), one is unrepresentative of the description (b), and the third is the conjunction of both simple targets (c). Tversky and Kahneman's hypothesis was that students would rank the compound target in between the two simple targets in terms of its probability. That hypothesis proved to be true: a large majority of students (85%) ranked probabilities according to this naïve hypothesis, labelled the 'psychological principle of similarity' by Tversky and Kahneman.

Following the research of Tversky and Kahneman, subsequent research aimed at the identification of a taxonomy of naïve theories or misconceptions, and at the development of assessment instruments. Empirical studies based on these kind of instruments, and our second chapter contains such a study, demonstrate that the level of misconceptions in students is unrelated, or at most very weakly related, to the level of knowledge of students. Although somewhat counterintuitive, this phenomenon is not at odds with the outcomes of studies in the coin problem and related science problems: mastery of formal scientific theories can go together with applying naïve theories when confronted with a problem outside the context of a course exam. And the phenomenon is not restricted to these two examples, or to science and statistics: the 'power of life over schooled knowledge' (Alexander, 2006) appears to be quite a strong empirical law. But if the state of true prior knowledge (that is: formal,

correct knowledge) cannot help us understand the state of naïve knowledge, what else can? Outside prior knowledge, learning theories based on learner-centred psychological principles distinguishes two broad areas of psychological theory that might help to explain the state of misconceptions: cognitive and metacognitive factors, and motivational and affective factors (Alexander; Alexander & Murphy, 2006). The role of both types of factors in the explanation of the states of naïve and formal knowledge has been investigated; in this thesis, we will focus however on the role of motivational and affective factors.

This thesis contributes to the empirical understanding of student learning. This is done by estimating models that represent elements of learning processes. The models are based on the integration of educational and psychological theories, such as expectancy-value models of achievement motivation and trait theory of personality. Besides these content theories, the estimation of learning models requires methodological guiding principles. The main modelling principle that will be used in this research, is based on the ‘presage-process-product model of classroom learning’ (Biggs 1993, 1999; Prosser & Trigwell 1999). According to this 3P modelling approach, learning is seen as a progression from presage (learning context) through process (learning acts) to products (learning achievement). The approach distinguishes several building blocks in explaining learning outcomes (see Figure 1.1, adapted from Prosser & Trigwell, 1999).

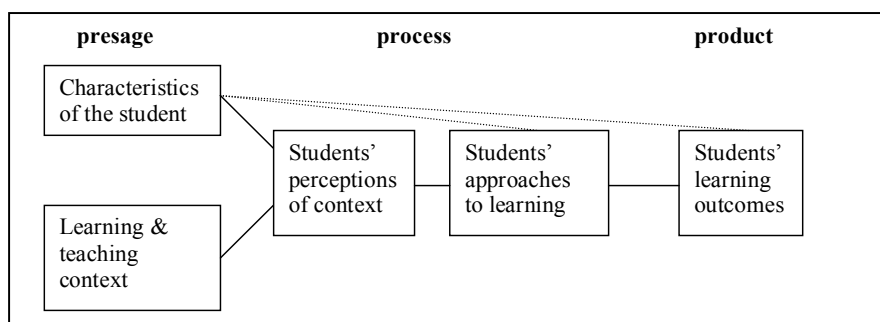


Figure 1-1. 3P or Presage-Process-Product Model of Classroom Learning

The student-based building block that is part of presage contains factors that explain learning from individual differences psychology. These factors include abilities, prior knowledge, motivation, personality, learning styles,

stabilised learning approaches, and so on. The second presage building block builds on traditional staff-developmental models, and focuses on teacher behaviour and the role of the learning environment. The process part of the model starts with the interaction of student characteristics and learning environment: this interaction determines the students' perceptions, the first process building block. Perceptions drive variables in the second process building block: the strategies selected for handling the task. The last building block, product, describes the outcomes achieved in the learning process. The complete model integrates teaching-based, student-based and process-based approaches to learning, and in addition to that, allows for feedback from process to presage, and from product to process and presage. Crawford et al (1998) constitute examples of the application of the 3P modelling approach to learning mathematics.

In addition to the 3P model, we will consider in this thesis the multilayered structure in personality characteristics, often expressed in the format of the 'onion metaphor' (Bakx, et al., 2006; Curry, 1983; Vermetten, Lodewijks, & Vermunt, 2001). According to this metaphor, several layers of personality constructs are distinguished, where inner layers are assumed to explain the more outer layers. The inner layers correspond to stable, trait-like characteristics, like the personality traits, whereas outer layers are less stable, and under direct influence of the environment. This onion metaphor is necessary since the 3P student-based building block contains more than one layer, and therefore brings about the need of a second organizing principle. The concept of a multilayered personality structure, suggesting that personality traits are more stable constructs than achievement motivations and should thus be regarded as causal, satisfies this need.

The estimation of models of learning behaviour in this thesis requires the identification of variables that are, at best, partly observed. This problem is referred to in the literature as the 'iceberg metaphor' (Bakx, et al., 2006; see also Spencer & Spencer, 1993). This metaphor does not have consequences for the model identification process, but much more for the model estimation process. The metaphor states that in learning processes, most of the relevant variables are hidden from observation, and only a very limited set of variables, mostly outcomes of learning processes, can be directly observed. Even then, one can wonder if learning outcomes as measured by e.g. grades on a final exam, are more than an approximation of the true knowledge state of the student. For all other variables in the model, this is certainly not an open question. Other variables belong to the 'under water level' part of the iceberg, cannot be observed in any direct

way, and can only be approximated plus or minus a measurement error. This metaphor illustrates the necessity of an 'errors in variables' approach, and the need of structural equation modelling tools as standard methodology. A structural equation model is distinct from a path model in that it hypothesizes that crucial variables, such as achievement motivations in this study, are not directly observable and are better modelled as latent variables. In doing so, SEM makes it possible to distinguish two different types of errors: errors of prediction in structural equations, and errors of measurement in the observation of variables. Models that involve only observed variables, as regression and path models, assume that measured variables are perfectly valid and reliable. In applications like this one, where all measurements take place through self-report questionnaires, the assumption of perfect validity and reliability is clearly unrealistic. Measured scales of our psychological constructs contain a measurement error component of an unknown size that is likely to vary over different instruments. If measurement error is present but not accounted for in the model estimation step, this might have serious consequences (Schumacker & Lomax, 2004). Therefore, all constructs under study are assumed to contain measurement errors, making the SEM approach appropriate. The SEM approach buys its methodological advantages at one, big, cost: it is very data-intensive. To this purpose, we have collected data on student learning of nearly 3000 first year students economics and international business. The large size of first year courses also explains our preference for quantitative research, whereas most empirical research in statistical reasoning is qualitative in nature, based on thinking-aloud sessions or student interviews (Garfield and Ben-Zvi, 2004).

Chapters 2 and 3 focus predominantly on the properties of statistical reasoning. Adopting the most widely known assessment instrument for measuring the presence of naïve statistical theories, the Statistical Reasoning Assessment or SRA of Garfield (2003, 1998a), these two chapters investigate the presence of naïve theories and the relationship without student background factors. The instrument is based on students' solutions of a set of statistical and probabilistic problems. Alternative solutions provided with these problems correspond to eight different types of misconceptions - when choosing an inappropriate solution - and eight related types correct statistical reason - when choosing the proper answer. Empirical studies based on the SRA instrument, including our research documented in chapter 2, demonstrate that the level of misconceptions in students is unrelated or, at most very weakly related, to the level of

knowledge of students, as measured by either course exams or entry tests. But if not related to knowledge factors, are there any other factors in the 3P model that might help to understand the development of naïve theories? Chapter 2 investigates the relationship of statistical reasoning to a wide range of factors relevant for the 3P model, whereas in chapter 3 we will focus on achievement motivations as potential predictor of statistical reasoning.

In Chapters 4 to 6 the main focus is on the demarcation between presage and process elements of student related factors. This focus is again nicely illustrated with the help of the 3P model. Figure 1 depicts the 3P model according insights of beginning and midst nineties. The building block “student characteristics” contains a broad range of affective and cognitive student related factors. In their application of the 3P model in the mathematical domain, Crawford et al. (1998) e.g. include prior conceptions of learning mathematics, prior conceptions of mathematics as subject, and prior orientations to study as part of presage factor. That contrasts assumptions at the basis of two recent UK projects investigating undergraduate learning processes. In the nationwide ETL project, or ‘Enhancing Teaching-Learning Environments in Undergraduate Courses’ (Entwistle, 1998, 2003; project website www.ed.ac.uk/etl), the search is for characterising strong teaching-learning environment in a range of undergraduate subjects areas taught at different studies and universities. And within the single institution of University of Oxford, the Oxford Learning Context Project explored undergraduate students’ perceptions of their learning environment, the ‘Oxford Tutorials’, in order to find ways of improving learning (Trigwell & Ashwin, 2003). In both UK projects, most explicitly in the Oxford Learning Context Project, Biggs’ 3P model was the starting point, and the project aimed to statistically specify the paths between all building blocks. Different from early studies based on the 3P model, Trigwell and Ashwin (2003) restrict the student related presage factors in their interpretation of the 3P model to an absolute minimum of ‘hard encoded’ biological and geographical factors: gender, year of study, prior schooling. Deviating from the original 3P model, they adapt the first process building block to incorporate all affective and cognitive student related factors, in combination the students’ perceptions of the environment. This adaptation can be understood as the product of the increased role of contextualisation: students do not develop motivation and conceptions in isolation, but in interaction with their environment, such as the relevant learning and teaching context. Contemporary insights on the

role of contextualisation do indeed suggest that student related affective and cognitive factors are influenced by context and students' conceptions of that context, but these insights seem not to be so outspoken to suggest that all of student related factors are context dependent. In other words: the black-and-white interpretations of student related affective and cognitive factors, corresponding to the original 3P model at the one side, and to the updated Oxford learning Context Project version of the 3P model at the other side, seem both unnecessarily and indefensibly extreme, and denying the interesting question about the true deepness of the greys. Although current 3P studies do not focus on this type of demarcation question, the decomposition of student related learning factors in a component unrelated to the relevant context, and a second component related to the context, is not new. Addressing the student factor self-concept, Marsh and co-authors have developed an extensive research tradition into the hierarchic, multidimensional nature of this important determinant of the learning process (Marsh, 1990; Marsh, Byrne, & Shavelson, 1988; Marsh et al., in press). The multi-dimensional nature is a consequence of self-concept being dependent upon the context: one can e.g. be very confident in doing math, but much less in doing sports, and vice versa. So subject domain is but one relevant aspect of context causing variation in self-concept, as will be teaching related factors. The variation of self-concept over different contexts does however still allow for the existence of a fixed, trait-like component: some students have high self-concepts in all contexts, others much lower. This component constitutes the hierarchic nature of self-concept. The assumption of self-concept being both hierarchic and multi-dimensional implies both components to be present, and allows the research question as to compare the relative size of both components. In a similar vein, Chapters 4 to 6 address that same question of the decomposition into a hierarchic and context-dependent components of achievement motivations based on the expectancy-value model. The hierarchic and multi-dimensional nature of achievement motivations is not at all self-evident. Laymen theories, and classroom practices based thereon, quite often regard a student either as motivated, or not, irrespective of subject domain. Moreover, motivation is often regarded as an amorphous characteristic that although it may contain different aspects, these aspects are empirically indistinguishable (Marsh & Yeung, 1996). Empirical research in achievement motivations seem to refute these laymen viewpoints: motivations are much more likely to be hierarchically ordered, multidimensional and multifaceted concepts (Alexander, 2006; Marsh &

Yeung, 1996; Marsh et al., 2006). The property of multifacetedness refers to the different aspects being much less amorphous than suggested: subject achievement motivations do contain different aspects that are empirically distinguishable and moreover might have very different characteristics over groups. A simple example of this stems from gender research: males tend to be overconfident in terms of self-perceived competence compared with females, but at the same time regard the domain to be more difficult than females do (Alexander). The property of multidimensionality refers to motivations being different for different domains: motivations tend to be subject domain specific. Gender research provides another example for this property: whereas male students tend to be relatively more motivated for math and science domains, female students are better motivated for the language related domains (Alexander). The existence of a hierarchy does restrict the domain specificity: achievement motivations can be decomposed into a component that is generic, and components that are subject specific. Most empirical research in these three properties of motivations has focussed on relative young students, taking middle or high school, and subjects domains being very distinct, such as languages, physical education, and math. In the Chapters 4 to 6 of this thesis, we address the same research question, but now brought to the level of higher education. Since programs in higher education contain subject domains that are more congruent than the middle and high school curriculum, this change in focus adds to the level of ambition of the research question, especially with regard to the property of multidimensionality. In Chapter 4, the hierarchical, multifaceted and multidimensional nature of achievement motivations is in itself the main topic of investigation. Having demonstrated that all three properties are present, Chapter 5 adds in the explanation of these properties: can individual differences with regard to personality factors, generally assumed to be one of the most stable exponents of personality, help in explaining individual differences in subject achievement motivations? Given the stability of personality factors and the variability of achievement motivations, causal relationship, if existent, must demonstrate variation over different subjects. Besides the characteristics of this variation of causal relationships over subjects, the strength of the relationships is an important issue: the stronger the relationship, the smaller the perspectives of any educational setting to improve students' achievement motivations. In Chapter 6, we will extend the discussion of the malleability of achievement motivations one step further by introducing an aspect that also played an

important role in Chapter 2: the genderedness of educational constructs. We will indeed succeed in demonstrating that both personality factors and subject specific achievement motivations are gendered, which gives rise to subsequent questions, as where to locate the origin of the gender-effect, in order to better understand the malleability of achievement motivations.

REFERENCES

- Alexander, P. A. (2006). *Psychology in learning and instruction*. Upper Saddle River, NJ: Pearson.
- Murphy, P. K. & Alexander, P. A. (2006). *Understanding how students learn*. Thousand Oaks, CA: Corwin Press.
- Bakx, A.W.E.A., van der Sanden, J.M.M., Sijtsma, K, Croon, M.A., & Vermetten, Y.J.M. (2006). The role of students' personality characteristics, self-perceived competence and learning conceptions in the acquisition and development of social communicative competence: A longitudinal study. *Higher Education*, 51, 71-104.
- Bernstein, P. L. (1996). *Against the gods: The remarkable story of risk*. Wiley.
- Biggs, J. B. (1993). From theory to practice: A cognitive systems approach. *Higher Education Research and Development*, 12, 73-86.
- Biggs, J. (1999). *Teaching for quality learning at university*. Buckingham: The Society for Research into Higher Education & Open University Press.
- Bransford, J. D. et al., (Eds.). (2000). *How people learn: Brain, mind, experience, and school. Expanded edition*. Committee on Developments in the Science of Learning with additional material from the Committee on Learning Research and Educational Practice, National Research Council, Washington: National Academy Press.
- Bruer, J. T. (1997). *Schools for thought: A science of learning in the classroom*. Cambridge, MA: MIT Press.
- Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, 50, 66-71.
- Crawford, K., Gordon, S., Nicholas, J. & Prosser, M. (1998). Qualitatively different experiences of learning mathematics at university. *Learning and Instruction*, 8, 455-68.
- Curry, L. (1983). *An organization of learning styles theory and constructs*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.
- Entwistle, N. (1998). *Concepts and conceptual frameworks underpinning the ETL Project*. ETL Project: Occasional Report 3.
- Entwistle, N. (2003). *University teaching-learning environments and their influences on student learning: An introduction to the ETL Project*. Paper presented at the EARLI conference, Padova.

- Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in statistics: Implications for research, *Journal for Research in Mathematics Education*, 19, 44-63.
- Garfield, J., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking*, (pp. 397-409). Dordrecht: Kluwer Academic Publishing.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision making under risk. *Econometrica*, 47, 263-291.
- Kahneman, D., & Tversky, A. (Eds.) (2000). *Choices, values, and frames*. New York, NY: Cambridge University Press.
- Marsh, H. W. (1990). A multidimensional, hierarchical self-concept: Theoretical and empirical justification. *Educational Psychology Review*, 2, 77-172.
- Marsh, H. W., Byrne, B. M., & Shavelson, R. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, 80, 366-380.
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality*, 74, 403-456.
- Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching. The experience in higher education*. Buckingham: The Society for Research into Higher Education & Open University Press.
- Marsh, H. W. & Yeung, A. S. (1996). The distinctiveness of affects in specific school subjects: An application of confirmatory factor analysis with the National Educational Longitudinal Study of 1988. *American Educational Research Journal*, 33, 665-689.
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.
- Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York, NY: Macmillan.
- Tversky, A. & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky, (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 84-100). Cambridge: Cambridge University Press.
- Tversky, A., & Shafir, E. (Eds.). (2004). *Preference, belief, and similarity*. Cambridge, MA: The MIT Press.
- Trigwell, K., & Ashwin, P. (2003). *Undergraduate students' experience of learning at the University of Oxford*. University of Oxford, Institute for the Advancement of University learning: Final Report of Oxford Learning Context Project.
- Vermetten, Y.J., Lodewijks, H.G., & Vermunt, J.D. (2001). The role of personality traits and goals orientations in strategy use. *Contemporary Educational Psychology*, 26, 149-170.

Chapter 2

PUZZLES IN STATISTICAL REASONING*

1. INTRODUCTION

The present study aims to explore the role of statistical reasoning in learning statistics. To this purpose the definition of statistical reasoning by Garfield and Chance (2000) will be employed. Statistical reasoning, according to their definition, is the way students reason with statistical ideas and make sense of statistical information. This involves making interpretations based on sets of data, representations of data, and statistical summaries of data. Statistical reasoning is based upon an understanding of important concepts such as distribution, location and variation, association, randomness, and sampling, and aims at making inferences and interpreting statistical results.

In order to measure statistical reasoning the Statistical Reasoning Assessment or SRA, developed by Garfield (1998a, 2003) has been used. Garfield and co-authors have performed several empirical analyses on the SRA (Garfield 1998b, 2003; Garfield and Chance 2000; Liu 1998). One of the striking outcomes of this research is the puzzle of 'non-existing relations with course performances': correlations between aggregated reasoning skills demonstrate low or zero correlations with course performances. A second puzzle emanating from this empirical work is the 'gender puzzle': female and male students demonstrate striking differences in their reasoning abilities. In addition to these two puzzles, a third, though less

* This chapter is based on: Tempelaar, D. T., Gijsselaers, W. H., & Schim van der Loeff, S. (2006), Puzzles in statistical reasoning. *Journal of Statistics Education*, 14(1).

surprising, effect is found: a country or nationality effect. This paper addresses these puzzles with the purpose to further the understanding of statistical reasoning and the assessment of it through the instrument SRA.

Statistical reasoning, and the related concepts of statistical thinking and statistical literacy, are at the center of interest of the educational statistics community. For example, the Winter 2002 edition of the *Journal of Statistics Education* provides a series of articles based on an American Educational Research Association (AERA) 2002 symposium: delMas (2002a), Garfield (2002), Chance (2002), Rumsey (2002) and delMas (2002b). The articles explore definitions, distinctions and similarities of statistical reasoning, thinking, and literacy, and discuss how these topics should be addressed in terms of learning outcomes for educational statistics courses. In the closing summary of the JSE Winter 2002 series, delMas (2002b) emphasizes the role of assessment. Although a lot of progress has been made in the delineation of the concepts reasoning, thinking and literacy, and the elaboration of instructional implications of research findings in each of the areas, we are still rather empty-handed with regard to instruments that assess students' abilities. In small-scale experimental settings, a range of techniques based on interviewing students, or think-aloud problem solving has been documented [see e.g. the contributions to the SRTL forums on Statistical Reasoning, Thinking and Literacy, of which the first two editions are reported in Ben-Zvi and Garfield (2004)]. Objective instruments that can be applied on a broad scale in classes as large as the one reported on in this study are, to our knowledge, limited to the SRA instrument.

The relationship between statistical reasoning (and related concepts) and the learning of statistics is a complex one. First of all, statistical reasoning is an achievement aimed for in most introductory statistics courses, comparable to traditional achievements as e.g. the understanding of the concept of sampling distributions. This is what Gal and Garfield (1997) call the outcome consideration. Expressed by Garfield (2002, p. 9): "it [statistical reasoning] appears to be universally accepted as a goal for students in statistics classes". But in addition to being an important output of statistics education, statistical reasoning is also a crucial input in the process of learning statistics: the process consideration. Students enter our classes with prior reasoning skills; to the extent that these prior skills correspond to true knowledge being part of the course achievements aimed at, these prior skills will ease the learning process. However, an important category of prior knowledge is formed by misconceptions, or intuitive but faulty reasoning mechanisms. Both types of preconceptions are, according

to modern learning theories (Bransford, Brown and Rodney 2000) crucial determinants in learning; if preconceptions are not properly addressed, newly learned knowledge might appear much more volatile than existing preconceptions brought into class. Research on learning in general (see e.g. Bransford, et al. 2000), and on statistical reasoning in particular (Garfield and Ahlgren 1988; Shaughnessy 1992), make clear that the intuitive misconceptions are of a stubborn nature. It has been demonstrated that even students who can correctly compute probabilities, tend to fall back to faulty reasoning misconceptions when asked to make an inference or judgment about an uncertain event outside the context of doing a statistics exam. They seem to rely on incorrect intuitions already present when entering the course. Therefore teaching correct conceptions - no matter how successfully - is no guarantee for students not applying misconceptions anymore. Examples of stubborn fallacies in student's statistical reasoning are the 'Law of small numbers' and the 'Representativeness misconception', both described in Kahneman, Slovic, and Tversky (1982), the 'Outcome orientation' described in Konold (1989), and the 'Equiprobability bias' described in Lecoutre (1992).

In the above mentioned studies, empirical analyses into statistical reasoning on the basis of the SRA-instrument has been performed by Garfield and co-authors. In all of these analyses the SRA was administered at the end of a course, parallel to the final exam. Their main aim was to investigate the mastery of reasoning skills and its relationship to course performances. In such a design, measured skills are a mixture of those newly achieved in the course, and those already present at the start of the course. In contrast to these studies, we administered the SRA in the very beginning of the first introductory course. Its outcomes are, thus, to be regarded as students' preconception levels achieved outside class or, in some cases, in high school programs, independent of our own curriculum. This difference in timing of administering SRA makes it possible to focus on the role of prior conceptions and misconceptions in learning statistics in our course.

Since reasoning abilities are measured in the very beginning of the course, instruction-related variables are excluded as a possible 'contaminant'. Differences in statistical reasoning can thus be wholly attributed to, what Garfield and Ben-Zvi (2004) call the 'diversity of students'. That diversity can refer to several student-related aspects. Ben-Zvi and Garfield (2004) contains a range of studies on statistical literacy, reasoning, and thinking in which student diversity expresses itself primarily

in differences in prior education and mastery. But student diversity has more manifestations than these cognitive aspects, and in this study, we will include two non-cognitive factors expected to have an impact on learning statistics. The first of these is constituted by the affective factor students' attitudes towards statistics. Attitudes are found to be important factors in learning statistics for several reasons: see for example Nasser (2004), Gal and Ginsburg (1994), Gal and Garfield (1997). Gal and Garfield (1997) distinguish e.g. between access considerations (the willingness of students to elect statistics courses), process considerations (their influence on the learning and teaching of statistics, the focus of this study), and outcome considerations (their role in influencing students' statistical behavior after leaving university). Analogous to its role in learning statistics, we hypothesise that positive attitudes contribute to a better state of prior reasoning abilities and misconceptions. A second aspect of students' diversity incorporated in this study is the typical way students tend to study: their learning strategies, or more generally, their learning approaches. In statistics education, this theme has received less attention than the role of attitudes, in contrast to empirical research in learning in general; see for example Biggs (2003), Bransford et al. (2000). Typically, learning theories based on student approaches to learning distinguish between deep and surface learning (Biggs 2003; Vermunt and Vermetten 2004; Duff, Boyle, Dunleavy, and Ferguson 2004). Students taking a deep learning approach are more or less our ideal students: triggered by an interest in the topic under study, these students focus on underlying meaning, on main ideas, principles and applications. In contrast, students taking a surface approach to learning are characterised by a focus on memorisation, rote learning, but no real attempt of understanding. In agreement with findings of general learning theory based on learning approaches, we hypothesize that a deep approach has a positive impact on, and a surface approach has a negative impact on, the level of statistical reasoning abilities students possess at the start of the course.

In Subsection 3.1 it is established that both puzzles and the nationality effect are indeed present in our data. The first puzzle, that of the non-existing relation between reasoning abilities and course performances, is studied in Subsection 3.2. In Subsection 3.3 we examine if part of gender and nationality effects can be retraced to differences in attitudes toward statistics on the basis of the hypothesis, made above that positive attitudes towards statistics contribute to a better state of prior reasoning abilities and misconceptions. In Subsection 3.4 we examine if part of gender effect can

be retraced to differences in learning approaches on the basis of the hypothesized different impacts on the level of statistical reasoning abilities that students possess at the start of the course. Integrating cognitive and non-cognitive factors explaining reasoning abilities, regression equations are presented in Subsection 3.5 for both reasoning abilities and misconceptions, where the role of gender appears to be restricted. It is not so much gender itself, but a complex of gendered characteristics describing a preferred learning approach of students that explains a limited but consistent part of statistical reasoning abilities. That integrative model addresses two of the puzzles and effects discussed in this contribution: the gender puzzle, and the nationality effect. Section 4 closes with discussion and educational implications.

2. METHOD

2.1 Setting and subjects of this study

The present study was conducted in the setting of a course Quantitative Methods (QM) being part of both Economics and Business first-year programs. It is an introductory course covering regular level-100 subjects from mathematics, statistics and computer skills. The mode of instruction of the course is one where students meet in small groups of approximately twelve students with a tutor to discuss their solutions to any - usually homework - problems supplemented by lectures.

Data were collected on three shifts of students: approximately 900 first-year students participating in the 99/00 QM course, and approximately 850 students participating both in 03/04, and in 04/05. In addition to those first-year students, another 10% percent of the students are 'repeat' students who did not manage to pass that specific course in previous years. All courses are taught in English. The faculty attracts a relatively large proportion of foreign students. In 99/00, the share of foreign students was 46%, a figure that has risen to 65% in 04/05. Of all foreign students, roughly two third has German nationality, the remainder being mostly from other European countries. Only the last couple of years, a growing but still rather small inflow of Asian students is visible. Distinguishing students according to nationality is important since major differences exist between secondary school systems in and outside Europe.

Most data used in this study are collected by students to be analyzed in their student projects. The topic of these projects has been 'a statistical

analysis of my study behavior,' in which the course participants compare their study habits with that of fellow students. In order to provide data for such a comparison, all students have completed several questionnaires in the first weeks of the course. The results, both individual data and aggregated group data, have been made available in the later weeks of the course. The SRA survey was one of the self-report instruments that students had to fill out in the first weeks of course. Other questionnaires that were administered are the Survey on Attitudes Towards Statistics (SATS,) and the Inventory of Learning Styles (ILS). The several questionnaires were administered in the tutorial sessions (99/00) or through web based forms (03/04 and 04/05). Due to the prospect of achieving bonus points for the student project, participation in the questionnaires was attractive and responses have been quite high. It is not possible to express the response rates as single figures, because different questionnaires were administered in different sessions (days), with different students being present. Most of the analyses reported here are based on the responses of about 2000 students (720, 580, and 700 in shifts 99/00, 03/04, and 04/05, respectively). The majority of the other students officially enrolled in the course would typically participate in the exam, but not in any educational activities.

2.2 Instruments: Statistical Reasoning Assessment, or SRA

18

The Statistical Reasoning Assessment, shortly SRA, is a multiple-choice test consisting of 20 items developed by Konold and Garfield as part of a project to evaluate the effectiveness of a new statistics curriculum in US high schools (Konold 1989; Garfield 1996, 1998a, 2003). In contrast to most other assessment instruments, it consists of closed format items, and it is therefore one of the few available instruments for large-scale assessment of statistical reasoning abilities of students at a pre-university level (see e.g. Gal and Garfield 1997, for a survey of assessment tools). Responses to items include a statement of reasoning, explaining the rationale for the particular choice. Some of these responses are instances of correct reasoning, but the majority demonstrate characteristic patterns of intuitive, incorrect reasoning. For a full description of the individual items and the eight correct reasoning scales and eight misconceptions scales, we refer to Garfield (1998a, 2003); Table 2.1 summarizes the several scales of the instrument.

Table 2-1. SRA Correct reasoning scales and misconceptions scales; based on Garfield (2003)

Correct Reasoning Scales:

CC1: Correctly interprets probabilities. Assesses the understanding and use of ideas of randomness, chance to make judgments about uncertain events.

CC2: Understands how to select an appropriate average. Assesses the understanding what measures of center tell about a data set, and which are best to use under different conditions.

CC3: Correctly computes probability, both understanding probabilities as ratios, and using combinatorial reasoning. Assesses the knowledge that in uncertain events not all outcomes are equally likely, and how to determine the likelihood of different events using an appropriate method.

CC4: Understands independence.

CC5: Understands sampling variability

CC6: Distinguishes between correlation and causation. Assesses the knowledge that a strong correlation between two variables does not mean that one causes the other.

CC7: Correctly interprets two-way tables. Assesses the knowledge how to judge and interpret a relationship between two variables, knowing how to examine and interpret a two way table.

CC8: Understands the importance of large samples. Assesses the knowledge of how samples are related to a population and what may be inferred from a sample; knowing that a larger, well chosen sample will more accurately represent a population; being cautious when making inferences made on small samples.

Misconception scales:

MC1: Misconceptions involving averages. This category includes the following pitfalls: averages are the most common number; failing to take outliers into consideration when computing the mean; comparing groups on their averages only; and confusing mean with median.

MC2: Outcome orientation. Students use an intuitive model of probability that lead them to make yes or no decisions about single events rather than looking at the series of events; see Konold (1989).

MC3: Good samples have to represent a high percentage of the population. Size of the sample and how it is chosen is not important, but it must represent a large part of the population to be a good sample.

MC4: Law of small numbers. Small samples best resemble the populations from which they are sampled, so are to be preferred over larger samples.

MC5: Representativeness misconception. In this misconception the likelihood of a sample is estimated on the basis how closely it resembles the population. Documented in Kahneman, Slovic, & Tversky (1982).

MC6: Correlation implies causation.

MC7: Equiprobability bias. Events of unequal chance tend to be viewed as equally likely; see Lecoutre (1992).

MC8: Groups can only be compared if they have the same size.

Studies reporting empirical data on the application of SRA are limited, and partly overlap in experiments they describe: Garfield (1998b, 2003), Garfield and Chance (2000), Liu (1998) and Sundre (2003). In an attempt to determine the criterion-validity of the SRA, Garfield administered the instrument to students at the end of an introductory statistics course and correlated their total correct and total incorrect scores with different course outcomes: final score, project score, quiz total (Garfield 1998b; Garfield and Chance 2000). The resulting correlations were low, suggesting that statistical reasoning and misconceptions were rather unrelated to students' performance in that first statistics course.

Garfield (1998b), Garfield and Chance (2000), and Liu (1998) report that the intercorrelations between items are quite low, implying a low reliability from an internal consistency point of view. In spite of these low intercorrelations, all of these studies analyze the total correct reasoning score and the total misconceptions score, so aggregated scores. The test-retest reliability for these two total scores turns out to be 0.7, and 0.75, respectively. We will follow the tradition of earlier studies in analyzing aggregated scores.

2.3 Instruments: Survey of Attitudes Toward Statistics, or SATS

Research in the affective domain of statistics education has lead to the development of several self-scoring instruments in the eighties, all using statements for which respondents mark their agreement or disagreement on 5-point or 7-point Likert-type; see Hilton, Schau, and Olsen (2004) for an overview. As each of these instruments had some drawbacks Schau, Stevens, Dauphinee, and DelVecchio (1995) developed the Survey of Attitudes Toward Statistics (SATS) in the nineties.

The SATS consists of 28 seven-point Likert-type items measuring four aspects of post-secondary students' statistics attitudes. The SATS contains four scales, see Schau, et al. (1995), Dauphinee, Schau and Stevens (1997), and Gal and Garfield (1997):

- Affect: measuring positive and negative feeling concerning statistics;
- Cognitive Competence: measuring attitudes about intellectual knowledge and skills when applied to statistics;
- Value: measuring attitudes about the usefulness, relevance, and worth of statistics in personal and professional life;

- Difficulty: measuring attitudes about the difficulty of statistics as a subject.

In a recent extension of the instrument, two more scales were added, each covered by four items: Interest, and Effort (better called planned effort, since the instrument is used as an *ex ante* measurement) (Schau 2004, personal communication). This extended version was available for the last of the three shifts of students incorporated in this study only. In our study, SATS was administered in the very first week of the course and can thus be viewed as an entry characteristic of the student.

2.4 Instruments: The Inventory of Learning Styles, or ILS

Students participating in our study made a profile of their own learning preferences using the instrument: Inventory of Learning Styles (ILS). The ILS aims at measuring the following components of student learning: cognitive processing strategies, metacognitive regulation strategies, conceptions of learning, and learning orientations (Vermunt and Vermetten, 2004; and numerous references in that source). The ILS consists of 120 statements covering all learning components. Students are asked to indicate, on a five-point scale, the degree to which they use the described learning activities in their studies, or to what degree the described views and motives correspond to their own views and motives. Table 2.2 describes the several scales within each of the learning components.

Table 2-2. Components and scales of the Inventory of Learning Styles

Processing strategies	Regulation strategies	Learning orientations	Mental models of learning
Relating and structuring	Self-regulation of learning processes	Personally interested	Construction of knowledge
Critical processing	Self-regulation of learning content	Certificate directed	Intake of knowledge
Memorising and rehearsing	External regulation of learning processes	Self test directed	Use of knowledge
Analysing	External regulation of learning results	Vocation directed	Stimulating education
Concrete processing	Lack of regulation	Ambivalent	Co-operation

2.5 Instruments: Course performance

In agreement with assessment literature (Gal and Garfield 1997; Jolliffe 1997), in our Quantitative Methods course learning performances are measured with a portfolio containing several instruments, each of them focusing on different aspects of the mastery of mathematical and statistical knowledge. Besides the before mentioned student project, the assessment instruments are:

- Final exams of the multiple choice format. To create a kind of external anchor, these exams are partly inspired by released Advanced Placement Statistics Exam. Like in the AP exam, our final exams will have a strong emphasis on conceptual issues, and students are allowed to use an extensive formula sheet, making the exam nearly of the 'open book' type. The exam covers both statistics and mathematics; both parts are graded separately.
- Quizzes of multiple choice and short answer format (in the 03/04 and 04/05 academic years and experimental in the 99/00 academic year). The quizzes allow students to achieve a bonus score. The level of the items is more basic than in the final exam, the main purpose being to stimulate student to spread their learning efforts evenly in time. It is hypothesized that the quiz score is stronger effort-based than the exam score.
- Weekly homework assignments of open type (only in the 99/00 academic year). The discussion of these assignments and the (partial) student solutions constitute the main agenda of the weekly, small-group, tutorial sessions. To get the discussions started, students were credited with some bonus for doing preparatory work on these assignments outside the tutorial group. Even more than the bonus for quizzes, these scores are assumed to be very strongly effort-based. Teaching assistants are explicitly instructed to assess the efforts put in by the students in trying to solve the homework problems, instead of assessing the correctness of the solution handed in. The success of the experiment with quizzes in the 99/00 shift led to the abandonment of the assessment of homework in later shifts.

3. RESULTS

3.1 Analysis of SRA data

Descriptive statistics of the present SRA data, similar to those reported in Garfield (1998b, 2003), Garfield and Chance (2000) and Liu (1998), are reported in Table 2.3. The exhibit presents the means of the several scales of all female and male students, and all Dutch and international students, expressed as a proportion, that is on a [0-1] scale. In addition to scores on eight reasoning skills (CC1 ... CC8), and eight misconceptions (MC1 ... MC8), the aggregated reasoning score (CCtot) and aggregated misconceptions (MCtot) are reported. The aggregated scores are obtained in the same way as in the studies by Garfield and co-authors by respectively taking the sum over all correct reasoning and misconception items, and re-expressing them as a proportion. Since the number of items per scale ranges from 1 to 5, different scales have a different weight in the total score, so aggregated scores are to be regarded as weighted averages. Added to the proportional scores are two measures that signal the existence of gender-effects and nationality-effects: the p-value of the independent samples t-test, and the Cohen's d measure of effect size, calculated as the difference in means divided by the pooled standard deviation (Cohen, 1988).

Table 2-3. Means of SRA Correct Reasoning scales and MisConceptions scales for Male and Female, Dutch and international students, and corresponding gender- and nationality effects

Correct Reasoning	Females		Males		p-value	Effect size	Dutch (N=1080)	Intern. (N=899)	p-value	Effect size
	(N=779)	(N=1209)	(N=779)	(N=1209)						
CCtot: total Correct Reasoning	0.54	0.57	0.54	0.57	.000	0.24	0.59	0.54	.000	0.44
CC1: Correctly interprets probabilities	0.70	0.72	0.70	0.72	.066	0.08	0.71	0.71	.929	0.00
CC2: Understands how to select an appropriate average	0.68	0.74	0.68	0.74	.000	0.24	0.77	0.68	.000	0.36
CC3: Correctly computes probability	0.38	0.43	0.38	0.43	.000	0.16	0.43	0.39	.003	0.13
CC4: Understands independence	0.63	0.60	0.63	0.60	.025	0.10	0.62	0.60	.199	0.06
CC5: Understands sampling variability	0.21	0.29	0.21	0.29	.000	0.27	0.30	0.23	.000	0.25
CC6: Distinguishes between correlation and causation	0.70	0.69	0.70	0.69	.490	0.03	0.78	0.62	.000	0.35
CC7: Correctly interprets two-way tables	0.71	0.78	0.71	0.78	.000	0.20	0.81	0.71	.000	0.27
CC8: Understands the importance of large sample	0.70	0.72	0.70	0.72	.205	0.06	0.73	0.70	.024	0.10
Misconceptions										
MCtot: total Misconceptions	0.33	0.30	0.33	0.30	.000	0.27	0.29	0.33	.000	0.34
MC1: Misconceptions involving averages	0.46	0.41	0.46	0.41	.000	0.21	0.38	0.48	.000	0.39
MC2: Outcome orientation	0.24	0.21	0.24	0.21	.001	0.15	0.23	0.22	.038	0.09
MC3: Good samples represent high % of population	0.17	0.14	0.17	0.14	.004	0.13	0.15	0.15	.566	0.03
MC4: Law of small numbers	0.33	0.25	0.33	0.25	.000	0.29	0.24	0.31	.000	0.24
MC5: Representativeness misconception	0.11	0.17	0.11	0.17	.000	0.22	0.15	0.14	.449	0.03
MC6: Correlation implies causation	0.24	0.26	0.24	0.26	.201	0.06	0.19	0.30	.000	0.33
MC7: Equiprobability bias	0.60	0.55	0.60	0.55	.001	0.16	0.56	0.59	.056	0.09
MC8: Groups can only be compared if same size	0.31	0.24	0.31	0.24	.002	0.14	0.27	0.27	.761	0.01

Outcomes in this and earlier studies are remarkably similar: Garfield (2003) e.g. reports as aggregate reasoning scores (CCtot) 0.56 and 0.60 for the U.S. and Taiwanese students, compared to 0.58 as the overall mean of CCtot in our study.

Similarity of our outcomes and those found in earlier studies is not limited to aggregated scores: also scale scores demonstrate very similar patterns. Of the correct reasoning scales, CC7 and CC8 are amongst those with highest mastery level, and CC3 and CC5 with lowest. Of the misconception scales, MC7 and MC8 are high in all studies (in our sample, MC8 somewhat less), and MC3, MC5 and MC6 are low.

In the Liu-study, reported in Garfield (1998b, 2003), Garfield and Chance (2000), and Liu (1998), the analysis of gender and country/nationality effects was restricted to the aggregated total correct and total misconceptions scores, instead of the individual scales. Based on an ANOVA of aggregated scores with country and gender as factors, Garfield (2003, p. 30) concludes: "It is interesting to see that despite the seemingly similar scale scores for the students in the two countries, there are actually striking differences when comparing the male and female groups. ... it will be interesting to see if replications of this study in other countries will yield similar results." 'Similar' should here be understood to mean that males have significantly higher total correct reasoning scores (except for the USA), and have significantly lower total misconceptions scores. These results can be generalized to our study with a remarkable regularity. We find significant gender effects in both aggregated scores in the same direction. Moreover, we find that CC2, CC3, CC5, and CC7 are significantly higher and MC1, MC3, MC4, MC7, and MC8 are significantly lower for males than for females among our students (where MC5 plays the role of the exception which proves the rule). All effects are quite strong in a statistical sense, having p-values below 0.005. The gender effect is rather substantial: males score more than 5% higher in total correct reasoning, and more than 9% lower in total misconceptions, than females, with Cohen's d effect size ranging between small and medium. Performing an ANOVA indicates that no interaction effects are present in our data; p-values of the interaction effect for CCtot and MCtot are e.g. 0.247 and 0.875, respectively. For that reason, no further ANOVA results are incorporated in this and subsequent subsections.

Conceptions for which we find higher scores than reported in the Garfield-studies, CC2, CC6, and CC7, may be characterized as general reasoning skills more than as statistical reasoning skills; higher 'European'

scores in general, and higher Dutch scores in particular, might simply reflect the general level of secondary education. Similar conclusions apply to the several misconception scales. We find high scores relative to the Garfield-reports for MC1, MC3, and MC6, all referring to topics that will be covered in any introductory course, so that the timing of the test administration might play a crucial role in explaining this difference (prior versus post assessment). In contrast, MC8 shows remarkably low misconception scores in our sample.

Similar to Garfield (2003), we find a nationality effect in half of all scales, and both aggregate scores. That effect has always the same direction: Dutch students have higher correct reasoning and lower misconception scores than foreign students. For both aggregated scales, Dutch students have an 11% higher total reasoning score, and a 9% lower misconception score, than non-Dutch students; effect sizes are in the range of medium. The nationality effect is about as stable as the gender effect, but much better explainable: Dutch secondary education seems to offer Dutch students a better preparation than most other European school systems, which shows up, amongst other things, in better general and statistical reasoning abilities. The focus on mathematics in Dutch secondary education, including an introduction into statistics and probability, which is rather uncommon in secondary school programs in other European countries, apparently provides Dutch students with a head start. Does this nationality effect possibly contribute to (part of) the gender effect? The answer is no; the female/male composition of Dutch and foreign student groups is very similar.

The second pattern refers to the high variability in prior mathematics education. Both Dutch students and students from most other European countries have taken mathematics in secondary school either as a major, or at advanced level, or alternatively as minor, or at basic level. Although the dummy 'math major' is a rather imprecise indicator of prior mathematics education, given the huge differences in mathematics programs in different European secondary school systems, it does contribute to the explanation of reasoning skills to a similar degree as nationality. Students with a math major have a 10.5% higher total correct reasoning score, and a 9.5% lower total misconception score, than students with a math minor. Apart from nationality, the math major dummy is a potential confounder explaining the gender puzzle since prior math education is somewhat biased, with 36% of the males versus 30% of the females having pursued a math major at high school level. However, the gender effect can only partially be contributed to

differences in prior math education. After splitting the sample into two sub samples, corresponding to different levels of prior math education, most scales still demonstrate significant gender effects.

As a last observation on average levels of reasoning skills and misconceptions, the high rate of correct answers is noticeable. Of the eight correct reasoning skills, five have means of above 65% correct. Of the eight misconception scales, only two have means larger than 35%. Given the circumstance that only a minority of our inflow did attend any formal education in statistics in secondary school, and a majority did not, one might doubt whether the level of the instrument is appropriate for (European) high schools and what impact the restricted discriminative power might have on the reliability of the instrument.

Correlations between the several SRA scale scores are low, and in many cases not significant. For correct reasoning skills, they range between -0.17 and +0.14, and for misconceptions, from -0.29 to +0.14. This finding is in line with other studies, see Garfield (1998b), Garfield and Chance (2000), Liu (1998), and Garfield (2003). As a consequence, the Cronbach alpha reliabilities of the aggregated scales, taking the eight correct reasoning scales and the eight misconception scales as components, are low: 0.29 and 0.11, respectively, and the focus on aggregated scales has therefore certain drawbacks. We will not pursue the issue of the reliability of aggregated scales here further, but will instead refer to Tempelaar (2004a, b) for alternative representations of the reasoning skills scales that avoid the reliability problems of aggregate scales.

3.2 SRA and student performance: the ‘non-existing relation puzzle’

3.2.1 Descriptive data about SRA and student performance

In this subsection we will focus on one of the three shifts of students: the 99/00 shift. Data of other shifts demonstrate similar patterns, but are less rich, since they lack one course performance instrument: the assessment of homework. The assessment portfolio that measures students’ course performance in the 99/00 shift contains three instruments: final tests, graded home work assignments and quizzes, each for mathematics and statistics, and each for three different periods. Descriptive analysis of the performance indicators shows, first of all, that the several performance indicators are strongly positively correlated. The strongest correlations are

amongst indicators of the same type. Correlations between final exam scores for math and stats in the three different periods range between 0.4 and 0.6; for homework assignments scores between 0.5 and 0.8, and for quizzes, even above 0.9. But correlations between scores of different types of assessment instruments are not much lower: between quiz scores and homework scores, ranging from 0.6 to 0.8, between quiz scores and final exam scores, ranging from 0.3 to 0.6, and between homework scores and final exam scores, ranging from 0.2 to 0.6.

Second: there exists a strong gender effect in both the quiz scores and bonuses achieved for homework assignments. This gender effect is present in mathematics and statistics, both for Dutch and international students, and always in the same direction: female students outperform male students. The effect is large, especially for the homework component. Third: there exists an even much stronger nationality effect in both performance indicators, where international students outperform Dutch students, both for mathematics and statistics, in all periods, both for females and males. Differences are again large.

With regard to the written exams, the picture is completely different. For all mathematics exams, and the first statistics exam, males outperform females, both for Dutch and for international students. In the second and third statistics exam, this pattern tends to reverse, females scoring higher than males; differences are however not significant. The nationality effect in exam scores demonstrates a somewhat similar development. In the first exam, Dutch students do significantly better than international students, both in math, showing a very large difference, and in statistics. In the second exam, Dutch and foreign students approach each other in math, whilst international students significantly outperform their Dutch counterparts in statistics. Finally, in the third exam, international students outperform Dutch ones both for math and for statistics significantly.

Most of these apparent differences have natural explanations. First of all the match between secondary education and university study is much better for Dutch students than for international students. The counter veiling force, though, is that international students, on average, put a lot more effort in their study than Dutch students. This difference in effort pays off in the more effort-based indicators such as bonus score for homework already from the very first period onwards, and starts to pay off in the more cognitive based indicators in the second period. The picture for the gender issue is similar: female students are willing to spend more efforts on their study than male students. This pays off starting from the very first period

onwards, especially in the effort-based bonus scores. However, it is not obvious why females start at a lower level in quizzes and exams, given the circumstance that differences in prior education are between small and absent.

3.2.2 SRA as predictor for performance indicators

What is the relationship between course performances and SRA scores, and how strong is this relationship? Regarding students' reasoning abilities as a relevant part of their prior knowledge base when entering the course, one would expect that correct conceptions would positively contribute to performance indicators, whereas misconceptions do the reverse, given that prior knowledge is in general one of the better predictors of course performance. Table 2.4 contains the correlations between aggregated SRA scales and performance indicators.

Table 2-4. Correlations of aggregated SRA scales and course performance indicators and their two-sided p-values: Homework bonus, scores in quizzes and final exam (N=680)

Performance indicator:	CCtot	p-value	MCtot	p-value
Homework bonus: Statistics period 1	-0.02	0.653	0.08	0.043
Homework bonus: Statistics period 2	-0.09	0.017	0.08	0.032
Homework bonus: Statistics period 3	-0.13	0.001	0.10	0.006
Homework bonus: Mathematics period 1	-0.12	0.001	0.14	0.000
Homework bonus: Mathematics period 2	-0.14	0.000	0.10	0.009
Homework bonus: Mathematics period 3	-0.06	0.093	0.03	0.388
Quiz score: Statistics period 1	0.01	0.792	0.00	0.922
Quiz score: Statistics period 2	-0.01	0.808	0.02	0.599
Final exam: Statistics period 1	0.24	0.000	-0.17	0.000
Final exam: Statistics period 2	0.06	0.131	-0.07	0.055
Final exam: Statistics period 3	0.07	0.072	-0.05	0.196
Final exam: Mathematics period 1	0.28	0.000	-0.18	0.000
Final exam: Mathematics period 2	0.18	0.000	-0.17	0.000
Final exam: Mathematics period 3	0.13	0.001	-0.17	0.000

Performance indicators are ranked such that they start in Table 4 with the most 'effort-based' indicators, the bonus for the weekly homework assignments, through the weekly quizzes, and finish with the least effort-based but strongly cognitive oriented final exams. This design is advantageous, because striking differences between the three assessment categories evolve. Starting with the written exams, we find a pattern that quite well fits the expectations: all significant correlations (and in fact, also

nearly all insignificant ones) between correct reasoning skills CCtot and performance indicators are positive and, although not very large, still substantial of size (up to .28). At the same time, all significant correlations with misconceptions are negative, but somewhat smaller in size. Weekly quizzes demonstrate a different pattern in that their relationship to SRA scales is absent. Going one step further into more effort-based indicators, the least intuitive result stems from the correlations between weekly homework bonus and SRA scales: all significant correlations have the 'wrong' sign, that is correct conceptions scores correlate consistently negative with bonus scores, and misconception scores correlate consistently positive with bonus scores!

This somewhat paradoxical result might explain why relationships between SRA scores and course performance can be weaker than the relationship between SRA scores and specific components of course performance. If the final course grade is composed as a weighted average of several assessment instruments, each of them having a different effort content, the aggregation process might cancel out the relationships between SRA scales and separate performance indicators. Alternatively, if progress tests like quizzes or mid term exams contribute strongly to grades, again a condition is created in which dependencies with SRA scales remain hidden. It is only through the two extremes, traditional final exams focusing on cognitive aspect on the one side, and scores for homework assignments on the other, that the impact of reasoning abilities and misconceptions becomes visible. In our analysis, we assume, as a working hypothesis, effort to be the mediating variable.

3.3 SRA and attitudes toward statistics

Except for Difficulty, students express positive attitudes towards statistics. This is true for all relevant subgroups of students; see Table 2.5. In contrast, mean scores for Difficulty are below the neutral level, expressing that students perceive the subject as difficult (the naming of the Difficulty-scale is somewhat counterintuitive: all scales are defined such that higher values correspond to more positive attitudes and feelings; a name like 'lack of perceived difficulty' would better catch this meaning).

Table 2-5. Average scores for SATS scales Affect, Cognitive Competence, Value and Difficulty, and the added scales Interest and Effort, and corresponding gender- and nationality effects, expressed by p -values and effect sizes

SATS Scales:	Females (N=822)	Males (N=1290)	p -value	Effect size	Dutch (N=987)	Intern. (N=1060)	p -value	Effect size
Affect	4.35	4.63	.000	0.28	4.70	4.37	.000	0.33
Cognitive Competence	4.79	5.06	.000	0.33	4.93	4.99	.123	0.07
Value	5.01	5.01	.848	0.01	4.97	5.07	.005	0.12
Difficulty	3.51	3.66	.000	0.20	3.76	3.46	.000	0.42

Added SATS Scales '04:	Females (N=276)	Males (N=439)	p -value	Effect size	Dutch (N=287)	Intern. (N=428)	p -value	Effect size
Interest	5.27	5.05	.002	0.24	4.94	5.27	.000	0.36
Effort	6.55	6.24	.000	0.44	6.08	6.55	.000	0.68

Table 2.5 indicates that both gender and nationality effects are present. Male students have significantly higher scores in Affect, Cognitive Competence and Difficulty, but significantly lower scores in Interest and Effort, than female students (all p -values being less than .005, and effect sizes ranging from small to medium); for Value, no significant difference exists. In comparing Dutch and international students, Dutch students express significantly higher Affect and Difficulty than international students, but lower Value, Interest and Effort; Cognitive Competence is invariant across nationalities (again at 0.005 level, with effect sizes ranging from medium to large). Attitude scores of our students are comparable to those reported in other studies; Schau (2003) e.g. reports pre-test scores for Affect, Cognitive Competence, Value, and Difficulty of 4.03, 4.91, 4.86, and 3.62, respectively.

Do attitudes as measured by SATS have any impact on students' state of reasoning abilities? If so, we expect this impact to be positive for the reasoning abilities, and negative for the misconceptions. The SATS instrument is based on the expectancy-value model of behavior, developed by Eccles and her colleagues (see, for example, Wigfield and Eccles, 2000, 2002; Eccles and Wigfield, 2002). According to this theory of achievement motivation, students' expectancies for success and the value they contribute to succeeding are important determinants of their motivation to perform achievement tasks. Expectation of success includes two components: belief about one's own ability in performing a task (the SATS scales Cognitive Competence), and a perception of the task demand

(Difficulty). From empirical research, these two aspects of success expectation are known to be positively related to the student's (prior) knowledge state (Wigfield and Eccles, 2000, 2002; Eccles and Wigfield, 2002). Therefore the expectation of positive correlations with reasoning, and negative with misconceptions, is most explicit for these two affects. These expectations turn out to be true, with the exception of the recently introduced variables Interest and Effort, as can be seen in the correlation matrix of Table 2.6.

Table 2-6. Correlations between SRA and SATS scales and their two-sided p-values (N=2031 for first four scales, N=687 for last two scales)

	CCtot	p-value	MCtot	p-value
Affect	0.12	0.000	-0.07	0.005
Cognitive Competence	0.12	0.000	-0.06	0.012
Value	0.10	0.000	-0.06	0.015
Difficulty	0.11	0.000	-0.10	0.000
Interest	0.02	0.533	0.05	0.174
Effort	-0.07	0.058	0.17	0.000

Although most correlations are very strongly significant, their size is moderate to small. In a joint analysis, SATS variables explain 2.2% of the variation in correct reasoning, and 4.5% in variation of misconception scores. However, the size of the gender effect is smaller, and since SATS variables are gender biased, the possibility of a gender effect induced by differences in attitudes is open.

By far the strongest correlation is the one between total MisConceptions and planned Effort in learning. This correlation is positive, a fact that contradicts the general hypothesis that positive attitudes will contribute to higher reasoning abilities and lower misconceptions levels, but it corroborates our working hypothesis formulated in the last section: learning approaches, characterized by investing large efforts, might result in inferior learning outcomes. However, there is another mechanism that has the potential to explain a positive relationship between the misconception level and planned effort: students realizing their deficient prior knowledge, might plan to compensate by spending above average efforts on their study. For this mechanism to apply, one would require a negative relationship between planned effort and prior knowledge. In our sample, we have three measurements that can be used to indicate prior knowledge: the SRA total reasoning abilities score, the students' score in math in the national exam (only for Dutch students), and most relevant, the students' self-scored

Cognitive Competence in the SATS instrument. Table 6 indicates that the correlation between SRA total reasoning score CCtot and Effort is absent. The same is true for the correlation between grade for the national exam, and planned effort. The third correlation, between Effort and Cognitive Competence, is significant, but its sign is opposite to what a compensation mechanism would predict. Higher self-concept is associated with higher planned efforts, thereby making the existence of a compensating mechanism very improbable, and in stead favoring the hypothesis of inadequate learning approaches.

3.4 SRA and student learning approaches

Analyzing the relationship between SRA and ILS produces correlations that are in line with other research into the relationship between learning approaches and course performance. Several significant correlations exist, but the size of them is restricted, typically being no larger than 0.1 (see e.g. Duff, et al. 2004). Deep processing typically contributes to better course performance, surface processing to inferior course performance. This pattern is also visible in our data on students' reasoning abilities: the deep processing component 'Critical processing' correlates positively to SRA correct reasoning and negatively to SRA misconceptions. The reverse is true for the surface component Analyzing. Table 2.7 contains all correlations and their significance levels.

The largest numbers of significant correlations are found amongst the last five scales, the mental models of learning. All these scores are positively related to the level of misconceptions, MCtot, and negatively associated with the level of correct conceptions, CCtot (except for Stimulating education). This finding deviates somewhat from the deep versus surface learning hypothesis; according to that hypothesis, one would expect that Construction of knowledge contributes to reasoning abilities, whereas Intake of knowledge would hinder it. From Table 7, one is inclined to adopt a different kind of mechanism; students with very outspoken mental models of learning (scoring high on one or two of the individual scales) tend to do worse in terms of reasoning abilities than students without outspoken mental models of learning who combine all or most of the individual models without being strongly dependent on any of them. A similar conclusion can be drawn from learning orientations, although the effect is weaker and restricted to Misconceptions. The learning orientations Vocation directed and Self test directed contribute positively to the MCtot score, as does Certificate directed, but with lower significance. This is to be

interpreted that a unidirectional learning orientation puts a student at a disadvantage in terms of misconceptions. Self-regulation of learning content and external regulation of the learning process have a, be it very modest, impact on Misconceptions of expected direction: students who do a better job in regulating their study themselves, achieve lower misconception scores. A similar impact on correct reasoning is absent. In general, it is noticeable that the strongest relations are those between learning approaches and misconceptions rather than between learning approaches and correct conceptions. In a joint analysis, learning approaches explain 5.1% of variation in MCtot, against 4.1% of variation in CCtot.

Table 2-7. Correlations between SRA and ILS scales and their two-sided p-values (N=1767)

ILS scale	CCtot	p-value	MCtot	p-value
Relating and structuring	0.03	0.242	0.02	0.366
Critical processing	0.10	0.000	-0.10	0.000
Memorizing and rehearsing	-0.02	0.466	0.01	0.687
Analyzing	-0.06	0.011	0.04	0.057
Concrete processing	-0.01	0.808	-0.01	0.799
Self-regulation of learning processes	0.00	0.894	0.00	0.832
Self-regulation of learning content	0.00	0.981	-0.07	0.005
External regulation of learning processes	-0.03	0.162	0.07	0.003
External regulation of learning results	0.00	0.977	0.04	0.058
Lack of regulation	-0.01	0.806	-0.04	0.107
Personally interested	-0.05	0.057	0.01	0.697
Certificate directed	-0.02	0.483	0.05	0.037
Self test directed	-0.04	0.064	0.11	0.000
Vocation directed	-0.04	0.125	0.12	0.000
Ambivalent	-0.06	0.016	-0.02	0.355
Construction of knowledge	-0.08	0.000	0.09	0.000
Intake of knowledge	-0.09	0.000	0.15	0.000
Use of knowledge	-0.05	0.034	0.11	0.000
Stimulating education	-0.02	0.336	0.07	0.004
Co-operation	-0.08	0.001	0.06	0.007

Can learning approach contribute to the explanation of the gender puzzle, and the corroboration of our effort hypothesis? The answer to both questions is affirmative. Correlations in Table 2.8 demonstrate that Effort is positively correlated with all four components of deep and surface processing. The strongest correlation is to be found between Analyzing and the SATS scale, where analyzing is the surface processing component

correlated with Misconceptions. The weakest correlation score can be observed for Critical processing, the deep processing component positively correlated with Correct reasoning. Finally, Effort is strongly positively correlated with all five learning orientations, each in their turn correlated with the Misconception score. Other attitudes are also correlated to learning approaches but, except for the two deep processing scales, these correlations are much weaker than those of the Effort scale.

Table 2-8. Correlations between selected ILS scales and SATS scale effort, and their two-sided p -values (N=675)

ILS scale	SATS Effort	p -value
Relating and structuring	0.22	0.000
Critical processing	0.12	0.002
Memorizing and rehearsing	0.19	0.000
Analyzing	0.29	0.000
Construction of knowledge	0.33	0.000
Intake of knowledge	0.22	0.000
Use of knowledge	0.31	0.000
Stimulating education	0.17	0.000
Co-operation	0.19	0.000

With regard to the gender effect, Table 2.9 contains the outcomes of tests on differences of means for the relevant ILS scales. The pattern is identical to that of Table 2.8: significant negative gender effects exist in scales that correlate strongly with the SATS Effort variable (Analyzing and all mental models of learning). In contrast, the deep learning component correlates most weakly with Effort, Critical processing, demonstrates the only significant positive gender effect.

Table 2-9. Gender effect (mean difference of males to females) in selected ILS scales, and two-sided p -values in an independent samples t -test (N=1215, 799 for males, females)

ILS scale	Gender effect	p -value
Relating and structuring	-0.008	0.810
Critical processing	0.100	0.001
Memorizing and rehearsing	0.013	0.659
Analyzing	-0.060	0.020
Construction of knowledge	-0.177	0.021
Intake of knowledge	-0.142	0.021
Use of knowledge	-0.142	0.024
Stimulating education	-0.131	0.024
Co-operation	-0.169	0.025

Motivated by the important role Effort appears to play in the growth of correct and incorrect statistical conceptions, we investigated in this section the relation between learning approaches and SATS scores. The global picture that emerges is that Critical processing, being an important constituent of the meaning-directed learning pattern, has a positive impact on statistical reasoning, whereas Analyzing, a constituent of the reproduction-directed learning pattern, has a negative impact. In addition, an outspoken mental model of learning and an outspoken learning orientation have negative impacts on statistical reasoning, whereas a more balanced mental model of learning and learning orientation contributes to better statistical reasoning. Since all these learning approach components appear to be gendered in our sample, they help explain the gender puzzle in statistical reasoning.

3.5 Final model and conclusions

Integrating the partial models of statistical reasoning explained by attitudes as well as learning approaches - including dummies for gender and nationality - generates regression equations as described in Table 2.10.

Explanatory variables	Response variables		CCtot: total Correct Reasoning score				MCtot: total MisConceptions score			
	Method Enter		Method Stepwise		Method Enter		Method Stepwise			
	β	t	β	t	β	t	β	t		
Intake of knowledge	-.017	-.499			.098***	2.886	.096***	3.538		
Use of knowledge	.008	.210			.042	1.142				
Stimulating education	.036	1.140			-.026	-.810				
Co-operation	-.044	-1.437			.011	.374				
Affect	-.041	-1.076			.062	1.625				
Cognitive Competence	.099**	2.553	.088***	3.223	-.032	-.831				
Value	.077***	2.676	.067**	2.462	-.075***	-2.623	-.063**	-2.464		
Difficulty	.011	.354			-.070**	-2.199	-.053**	-2.028		
R-squared	.091		.085		.088		.078			

Explained variation in the two regression equations achieved by stepwise regression is 8.5%, and 7.8%, respectively for Correct reasoning and MisConceptions. Nationality and gender dummies only explain 5.8% respectively 4.3% of variation, so adding both attitudes and learning approaches as predictors has a significant, but restricted effect on explained variation. The best predictor of Correct reasoning is the nationality dummy, contributing about half of all explained variation, followed by the learning approach Critical processing. The Gender dummy is significant, but has a very restricted impact: it explains less than 1%. Conclusions for the SRA MisConceptions variable are similar: nationality dummy and Critical processing are the main regressors, gender is significant with an impact stronger than in the correct reasoning case, but still limited. Reducing the sample to the 03/04 shift to allow the SATS variable Effort into the model has no effect in the first equation explaining Correct reasoning. It, however, has an impact on the second equation: Effort enters the equation replacing the gender dummy completely.

What can we conclude from this final model with regard to the research question of this study? First of all, there exists a solid nationality effect in both Correct conceptions and Misconceptions that overshadows all other effects. Although the nationality effect, in principal, can consist of several elements, the large differences between secondary school systems in European countries and the prominent role of statistics in the math program of Dutch high schools suggest that this effect is mainly caused by differences in prior schooling. The fact that the nationality effect is stronger in Correct reasoning than in MisConceptions, reinforces the plausibility of a schooling effect. Beyond the nationality effect, there exists a gender effect. However, SRA scales are not the only gendered phenomena relevant in statistics education; also attitudes toward statistics, as measured by SATS, and preferred learning approaches, as measured by ILS, demonstrate gendered components. For that reason, the greatest part of the gender effect in SRA (but not all of it) can be explained by differences in learning approaches and attitudes. Students with a reproduction directed learning pattern and unilateral learning orientations as well as mental models of learning are outperformed in statistical reasoning by students with a meaning-directed learning pattern along with balanced learning orientations and mental models of learning. Since female students are overrepresented in the first category (at least in our sample), a gender puzzle is created. A gender puzzle that arises most prominently in the MisConceptions part.

In doing this study, both puzzles seem to be -at least for the greatest part- resolved. What is left is the question of why statistical reasoning behaves so differently from other academic subjects, including mathematics and statistics. When confronted with a learning task, students will decide upon their preferred approach toward that task. That choice is first of all context dependent: students choose different learning approaches for different learning tasks. It is also student dependent: some students have 'on average' a stronger tendency to use surface approaches, others a stronger tendency to apply deep approaches. Empirical research in learning approaches generally indicates that although students with a stronger emphasis on deep approaches are somewhat more successful than students who emphasize surface approaches, approaches are best regarded as substitutes. There are different ways to reach the same goal, one maybe more efficient than the other, but in the end all resulting in mastery. The strong correlations found in this study between the several types of course performance, both for mathematics and for statistics, confirm this perspective. In this rather general pattern, statistical reasoning makes the exception. It's negative correlations with effort, and with several of the scales of the learning styles instrument, suggest that statistical reasoning calls for a unique learning approach, excluding alternative ways to mastery. 'Trying harder' has not many, but at least one limitation.

4. DISCUSSION AND EDUCATIONAL IMPLICATIONS

Most statistics programs, adapted to the education reform movement, contain a portfolio of different course assessments. Some assessment instruments are highly effort-based, as homework assignments and projects, while some are more cognitive based, such as final exams. In general, correlations between course outcomes as assessed by these different instruments tend to be rather high. Grading students with a portfolio, instead of a single final exam, thus seems not to have a strong impact on grading decisions. Choosing for a rich portfolio is therefore better understood by the desire to stimulate students in their learning, than to drastically change the grading outcomes.

The SRA-instrument is a natural candidate for any assessment portfolio in introductory statistics. However, in comparing its outcomes with other types of course performance, it takes a unique position: correlations with final exam outcomes are weakly positive, correlations with effort-based

instruments as homework assignments are however weak but negative. The weakness in the positive correlations found in this study might not be that problematic, though: it is after all a pre-test, and reasoning skills as measured by SRA are not included explicitly as course goals. More problematic might be the negative (be it weak) relationship between study efforts (as measured by the bonus for homework assignments) and the SRA outcomes. One interpretation of this is that a learning approach that is reproduction directed and strongly effort-based might be an obstacle in developing statistical reasoning. If this interpretation is correct, it will have a strong impact on statistics education. The assessment portfolio relevant for this study demonstrates a wide range of instruments: from multiple choice final exams, via quizzes, to assessed home work. Still, for all these different instruments, both deep and surface learning approaches contribute to achieving satisfying outcomes. So, although an effort focused learning approach might be not the most efficient to pass the course, it still carries the guarantee for success, as long as effort levels are high enough. If the SRA were to be added to the portfolio of assessment instruments, this story would become different. If our sample is representative, and if the characteristics of the SRA as post-test are similar to those of a pre-test, we cannot but conclude that there are no alternative routes toward achieving reasoning abilities.

One of Garfield's (2002) conclusions is that the quality of teaching, and the performance of students on their exams, does not tell that much about students' reasoning skills and their level of integrated understanding. This study adds to that that also specific aspects of the quality of learning, such as approaching learning tasks in a committed but reproduction directed way, do not guarantee proper reasoning skills. Chance (2002) describes several instructional tools that allow 'thinking beyond the textbook'. The outcomes of this study emphasize the importance of using those types of activities and other tools discussed by Chance; neither traditional lecturing, nor textbook-based independent learning, can assure success. The study at the same time indicates what those tools should do beyond teaching some specific skills or knowledge: strengthen e.g. critical processing, and create a better balance in learning orientations and mental models of learning, since these are important in achieving statistical reasoning skills.

REFERENCES

- Ben-Zvi, D., & Garfield, J. (Eds.). (2004). *The challenge of developing statistical literacy, reasoning, and thinking*, Dordrecht: Kluwer Academic Publishers.
- Biggs, J. (2003). *Teaching for quality learning at university* (2nd ed.). Buckingham: Society for Research into Higher Education / Open University Press.
- Bransford, J. D., Brown, A. L., & Rodney, R. C. (Eds.). (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. Committee on Developments in the Science of Learning with additional material from the Committee on Learning Research and Educational Practice. National Research Council, Washington: National Academy Press.
- Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education* [Online], 10 (3). www.amstat.org/publications/jse/v10n3/chance.html
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dauphinee, T. L., Schau, C., & Stevens, J. J. (1997). Survey of Attitudes Toward Statistics: Factor structure and factorial invariance for women and men. *Structural Equation Modeling: A Multidisciplinary Journal*, 4 (2), 129-141.
- delMas, R. C. (2002a). Statistical literacy, reasoning, and learning. *Journal of Statistics Education* [Online], 10 (3). www.amstat.org/publications/jse/v10n3/delmas_intro.html
- delMas, R. C. (2002b). Statistical literacy, reasoning, and learning: A commentary. *Journal of Statistics Education* [Online], 10 (3). www.amstat.org/publications/jse/v10n3/delmas_discussion.html
- Duff, A., Boyle, E., Dunleavy, K., & Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and Individual Differences*, 36, 1907-1920.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109-132.
- Gal, I. & Garfield, J. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal & J. Garfield (Eds.), *The Assessment Challenge in Statistical Education* (pp. 1-13). Voorburg: IOS Press.
- Gal, I. & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education* [Online], 2 (2). www.amstat.org/publications/jse/v2n2/gal.html
- Garfield, J. (1996). Assessing student learning in the context of evaluating a chance course. *Communications in statistics; Part A: Theory and Methods*, 25 (11), 2863-2873.
- Garfield, J. (1998a). *Challenges in assessing statistical reasoning*. AERA Annual Meeting presentation, San Diego.
- Garfield, J. (1998b). The Statistical Reasoning Assessment: Development and validation of a research tool. In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, &

- W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 781-786). Singapore: International Statistical Institute.
- Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education* [Online], 10(3).
www.amstat.org/publications/jse/v10n3/garfield.html
- Garfield, J. (2003). Assessing statistical reasoning. *Statistics Education Research Journal* [Online], 2(1), 22-38.
[http://www.stat.auckland.ac.nz/~iase/serj/SERJ2\(1\).pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ2(1).pdf)
- Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in statistics: Implications for research. *Journal for Research in Mathematics Education*, 19, 44-63.
- Garfield, J., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 397-409). Dordrecht: Kluwer Academic Publishers.
- Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematics Thinking and Learning*, 2(1&2), 99-125.
- Hilton, S. C., Schau, C., & Olsen, J. A. (2004). Survey of Attitudes Toward Statistics: Factor structure invariance by gender and by administration time. *Structural Equation Modeling*, 11(1), 92-109.
- Jolliffe, F. (1997). Issues in constructing assessment instruments for the classroom. In J. Garfield & I. Gal (Eds.), *The assessment challenge in statistical education*. Voorburg: IOS Press.
- Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6, 59-98.
- Lecoutre, M.P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics*, 23, 557-568.
- Liu, H.J. (1998). *A cross-cultural study of sex differences in statistical reasoning for college students in Taiwan and the United States*. Doctoral dissertation, University of Minnesota, Minneapolis.
- Nasser, F. M. (2004). Structural model of the effects of cognitive and affective factors on the achievement of Arabic-speaking pre-service teachers in introductory statistics. *Journal of Statistics Education* [Online], 12(1).
www.amstat.org/publications/jse/v12n1/nasser.html
- Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education* [Online], 10(3).
www.amstat.org/publications/jse/v10n3/rumsey2.html
- Schau, C. (2003). Students' attitudes: The "other" important outcome in statistics education. Paper presented in the Joint Statistical Meetings, San Francisco, CA.
- Schau, C., Stevens, J., Dauphinee, T. L., & Vecchio, A. De (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement*, 55(5), 868-875.

- SERJ (2002): *Statistics Education Research Journal*, 1 (1), 30-45. The International Research Forums on Statistical Reasoning, Thinking and Literacy: Summaries of presentations at SRTL-2.
- Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 465-494). New York: Macmillan.
- Sundre, D. L. (2003). *Assessment of quantitative reasoning to enhance educational quality*. AERA annual meeting presentation, Chicago. Available through the ARTIST web site:
www.gen.umn.edu/artist/articles/AERA_2003_QRQ.pdf.
- Tempelaar, D. (2004a). Statistical Reasoning Assessment: an analysis of the SRA instrument. *Proceedings of the ARTIST Roundtable Conference on Assessment in Statistics*. [Online]
www.rossmanchance.com/artist/proceedings/tempelaar.pdf
- Tempelaar, D. (2004b). Statistical Reasoning Assessment: an analysis of the SRA instrument. In *2004 ASA Proceedings of the Joint Statistical Meetings* (pp. 2797-2804). Alexandria, VA: American Statistical Association.
- Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review*, 16(4), 359-384.
- Wigfield, A., & Eccles, J.S. (2000). Expectancy - value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68-81.
- Wigfield, A., & Eccles, J.S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Eds.), *Development of Achievement Motivation* (pp. 92-120). San Diego: Academic Press.

Chapter 3

ATTITUDES TOWARD STATISTICS AND PRIOR REASONING ABILITIES*

1. INTRODUCTION

Recent research into statistical reasoning about variation, statistical reasoning about distribution, and statistical reasoning about sampling distributions, has created important insights into the developmental process of statistical reasoning skills. Most research has focused on the identification of subsequent, hierarchically ordered stages of reasoning development by means of qualitative research methods such as thinking-aloud sessions or interviews. Two recent issues of this journal, *Statistics Education Research Journal* (2004:2; 2005:1) and an edited volume (Ben-Zvi & Garfield, 2005), contain a wealth of empirical studies of this type into the cognitive process of developing reasoning abilities, and instructional tools that might foster these developments. The present research investigates statistical reasoning from a somewhat different perspective. It examines individual differences between students learning statistics and statistical reasoning. These individual differences demonstrate strong variability: students enter learning processes with different background characteristics, and different perceptions of the learning context. In as far as students' learning paths are dependent upon individual differences, this will give rise to diversity in learning paths. As a manifestation of students'

* This chapter is based on: Tempelaar, D. T., Schim van der Loeff, S., & Gijssels, W. H., A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities and course performance. Manuscript under revision for publication in *Statistics Education Research Journal*.

heterogeneity, this study uses students' prior attitudes toward statistics as measured by the extended Survey of Attitudes Toward Statistics (SATS) instrument (see Schau, Stevens, Dauphinee & DeVecchio 1995; Schau, personal communication, November 30, 2003). The main aim of this contribution is to investigate the relationship between students' prior statistical reasoning abilities when entering an introductory statistics course and their attitudes toward statistics. We identify statistical reasoning by the Statistical Reasoning Assessment (SRA) instrument (see Garfield 1996, 1998a, 2003).

Contemporary research in statistics education distinguishes an array of different but related cognitive processes in learning statistics: statistical literacy, statistical reasoning, and statistical thinking: see e.g. the special issues of *Journal of Statistics Education*, (2002:3), the two special issues of *Statistics Education Research Journal* (2004:2; 2005:1) and Ben-Zvi & Garfield (2005a). The demarcation of these three cognitive processes not being complete, it is well accepted that statistical literacy represents the most basic skills (Ben-Zvi & Garfield, 2004). Gal (2004) distinguishes two interrelated components in statistical literacy: the ability to "interpret and critically evaluate statistical information, data-related arguments, and stochastic phenomena", and the ability to "discuss or communicate" these (see also Rumsey, 2002). Statistical reasoning is the ability to "explain why a particular result is expected or has occurred, or explain why it is appropriate to select a particular model or representation" (DelMas, 2004; see also Garfield & Chance, 2000; Garfield, 2002). Statistical thinking involves an "understanding of why and how statistical investigations are conducted and the 'big ideas' that underlie statistical investigations" (Ben-Zvi & Garfield, 2004; see also Pfannkuch & Wild, 2004; Chance, 2002). Literacy, reasoning, and thinking are to some extent achieved even before formal schooling in statistics takes place. Those naive conceptions learned outside school can both be correct and incorrect in nature. In the 1970s, cognitive research into statistical and probabilistic reasoning, has revealed several categories of fallacies in human's reasoning, with examples as the 'Law of small numbers', the 'Representativeness misconception' (Kahneman, Slovic, & Tversky, 1982), the 'Outcome orientation' (Konold, 1989), and the 'Equiprobability bias' (Lecoutre, 1992). Most of that research is documented in the seminal work of Kahneman, Slovic, and Tversky (1982) (Garfield & Ahlgren, 1988). In the decades thereafter, following the reform movement in statistics education, research shifted its focus from probabilistic reasoning to reasoning with data (Pfannkuch & Wild, 2004), as

evidenced in the topics of the recent series of SRTL research forums and the compilation of their major contributions in Ben-Zvi and Garfield (2004). Another important development in recent decades is the design of assessment instruments for statistical literacy, reasoning, and thinking (DelMas, 2002; Garfield & Ben-Zvi, 2005b). Paraphrasing Chance (2002), 'if not assessed, it cannot be valuable'. Several instruments grew out of the need for assessment in the context of research projects. Characterizing these instruments, Garfield and Ben-Zvi (2004, p. 399) ascertain: "It may seem strange, given the quantitative nature of statistics, that most of the studies ... include analyses of qualitative data, particularly videotaped observations or interviews." Quantitative instruments are still scarce, and are all derived from the first and most prominent instrument in the field: Statistical Reasoning Assessment (SRA). The SRA was developed by Konold and Garfield (Konold, 1989; Garfield, 1996, 1998a, 2003) as part of a project evaluating the effectiveness of a new statistics curriculum in US high schools. The instrument is based on the well-described classes of misconceptions and their antipodes, the learned or unlearned correct conceptions, that emerged from cognitive science research into reasoning fallacies (Garfield, 2003; Garfield & Ahlgren, 1988). In current terminology - the SRA was developed long before recent discussions on the demarcation of literacy, reasoning, and thinking - fallacies addressed in the SRA are of all three types. Being designed in the earlier stages of the reform movement in statistics education (Ben-Zvi & Garfield, 2004), the SRA focuses both on statistical and probabilistic reasoning. Newer assessment instruments, related to the SRA but focusing stronger on reasoning with data, are currently being developed in the framework of the Assessment Resource Tools for Improving Statistical Thinking (ARTIST) project (DelMas, 2004b; see also <https://app.gen.umn.edu/artist/>). As long as these new instruments are not yet available, the SRA is certainly the most appropriate tool to assess students' reasoning abilities in large scale applications typical for educational practice.

Empirical studies on statistical reasoning focus predominantly on the cognitive developmental process students go through when learning reasoning abilities, and instructional tools that may foster these developments. The large majority of these studies is empirical in nature in that they use descriptions, often achieved by thinking-aloud sessions, or interviews of the cognitive states of students, to reconstruct a developmental trajectory (Ben-Zvi & Garfield, 2004a). Different states of students' reasoning abilities allow identifying subsequent stages in the

developmental process. Our study chooses the different perspective based on individual differences in student-related factors by investigating the role of non-cognitive individual differences in the cognitive development of students. This type of study has, at least in the context of statistics and mathematics education, a long tradition (Gal & Garfield, 1997; McLeod, 1992). In conceptualizing the non-cognitive domains of education, McLeod (1992) distinguishes between emotions, attitudes and beliefs. In most studies of learning processes in statistical education, the focus is on beliefs and attitudes, rather than emotions; see e.g. Gal and Ginsburg (1994), Gal and Garfield (1997). Probably the best known, and certainly most validated, model on the role of attitudes in learning statistics is the model developed by Schau and co-authors (Schau, Stevens, Dauphinee and DeVecchio, 1995). The Schau-model is based on the expectancy-value model for achievement motivations designed by Eccles and Wigfield (Eccles & Wigfield, 2002; Wigfield & Eccles, 2000, 2002). In that model, students' expectancies for success and the value they contribute to succeeding are important determinants of their motivation to perform achievement tasks. Expectancy for success crystallizes in two different concepts: belief in one's own ability to perform a task, and a perception of the task demand. Subjective task value is generally modeled in a single concept, comprising several aspects: attainment value (importance of doing well on a task), intrinsic value (interest in and enjoyment gained from doing the task), utility value (usefulness), and costs (spent efforts) (Eccles, 2005). The contribution of Schau and co-authors to the development of the expectancy-value model of achievement motivations is twofold. First, they designed a measurement instrument, the Survey of Attitudes Toward Statistics (SATS), to adapt the generic expectancy-value model to the statistical domain (Schau et al., 1995; Dauphinee et al., 1997). Second, they extended the generic model by introducing new concepts obtained by disentangling the broad task-value concept of the expectancy-value model. In the first 28-item version of SATS, the task-value concept is broken up into an affective concept, focusing most on the enjoyment aspect of intrinsic values, and a valuation concept, focusing on the remaining components of attainment and utility values. The model of the first version thus contains two expectancy factors that deal with students' beliefs about their own ability and perceived task difficulty: Cognitive Competence and Difficulty, and two subjective task-value concepts that encompass students' feelings toward and attitudes about the value of the subject: Affect and Value (Schau, 2003). Empirical research, both within the domain statistics (Dauphinee et al., 1997; Sorge &

Schau, 2002; Hilton et al., 2004) and in other academic domains (Tempelaar et al., 2007a) support the distinction of these affective and valuation aspects. In a second, 36-item version of SATS (Schau, personal communication, November 30, 2003), two more concepts are introduced: Interest and Effort. The Interest concept shapes the interest aspect of the intrinsic value component in the expectancy-value model, whereas the Effort concept shapes the perceived costs component in the subjective task-value (Eccles, 2005). To the knowledge of the authors, no empirical studies based on the extended SATS instrument have yet been published. Empirical studies of the 28 item version of SATS, referred to above, focus on the structure of attitudes alone, or on the structure of attitudes in relation to statistics course performances. The context of these studies is thereby slightly different from most studies in the expectancy-value framework that focus primarily on the relation between attitudes and learning task choices (such as course selection) rather than learning task outcomes.

The main contribution of this paper is to investigate the dependency on their attitudes toward statistics of students' prior reasoning abilities when entering an introductory statistics course. In the formulation of this research question, attitudes are hypothesized to be causal to statistical reasoning abilities. The hypothesized direction of causality is in agreement with process models of learning, see e.g. Garfield et al. (2002), in which affective, student related factors are regarded as determinants for cognitive, learning outcome related factors. In addition, attitudinal variables possess a trait-like nature, in contrast to reasoning abilities that possess a state-like nature. Therefore, the hypothesized causal direction follows the general modeling pattern of stable traits determining malleable states. In order to do so we will start the empirical third section by developing confirmatory latent factor models for attitudes, based on the extended SATS instrument, and for statistical reasoning, based on the SRA instrument. Subsequently, these factor models are integrated into a full structural equation model that explains reasoning abilities by attitude factors. To be able to put this relationship into perspective, two further cognitive constructs are added to this model: course performance measured by quiz and final exam scores. This extension allows characterizing reasoning abilities not only by their direct relationship with attitudinal variables, but also by a comparison of that relationship with the ones between attitudes and course performances.

One of the implications of our model is that where different learning approaches provide alternative routes to achieve traditional course performances, the one maybe more efficient than the other but all

contributing to the same learning goal, this seems not to be true for statistical reasoning abilities: some learning approaches really hinder achieving reasoning skills. The model outcomes thus have strong implications for the development of instructional programs in statistical reasoning, which will be one of the topics discussed in the concluding section.

2. METHOD

2.1 Participants and procedure

In this study, statistical reasoning was investigated of students participating in the programs “International Business” and “International Economics” of the Maastricht University. 842 respectively 776 students in these two programs participated in 04/05 and 05/06 in the first year, first semester course Quantitative Methods (QM). This is a compulsory introduction to mathematics and statistics for all students. Of these 1618 students, 64% are male, and 36% are female. Another relevant decomposition is: 39% students have a Dutch secondary school diploma, against 61% students with non-Dutch diplomas (most of them of German nationality).

Part of the data analyzed in this study comes from regular student examinations. In the QM course, three assessment instruments are applied. The first is a final, written exam, in multiple-choice format, covering both statistics and math. Items in the exam focus on the ability to apply statistical and mathematical methods; those in statistics are motivated by the exams of the Advanced Placement Program Statistics. Both for statistics and mathematics, three quizzes are taken spread over the eight weeks of the course. Quizzes are optional; they give rise to bonus points for the exam score. In practice, all students participate in most of the quizzes. For this study, quiz scores are aggregated over the three quizzes. The third assessment instrument is a student project. For this project, students collect personal data by filling several self-report instruments. Later on, they perform an explorative data analysis of these data as a student project. Students are informed that the self-report data are used for three other purposes next to being part of the project: for purpose of study advice to students who have adopted an inefficient study approach, for course-improvement purposes, and for research. The project is compulsory, and

assessed with pass/fail. Since students can acquire feedback on their project in the several stages of its development, the final assessment of it is not very informative, and is not included in this study.

The SATS and the SRA were the two first self-report instruments to be administered during the first days of the course. Responses to both surveys therefore reflect students' prior attitudes and beliefs toward statistics and their prior reasoning abilities. Scores cannot be influenced by impressions of the educational process, nor by knowledge achieved in the course itself.

Both instruments are quantitative of nature, and generate observations that can be regarded as proxies for the underlying, but unobservable theoretical constructs. Therefore, the investigation of the relationship between attitudes and reasoning abilities requires the estimation of two confirmatory latent factor models for attitudes on the one side, and for statistical reasoning on the other, and the integration of both these factor models into a full structural equation model. To this model, we add two indicators of course performance: latent variables measuring the strongly cognitive based scores in the final exam, and the more effort-based scores in quizzes. Prime reason to do so is that it allows characterizing the particular position statistical reasoning takes within the spectrum of different performance indicators.

2.2 Measures

Statistical reasoning abilities. The Statistical Reasoning Assessment (SRA), is a test consisting of 20 multiple-choice or multiple-answer items developed by Konold and Garfield as part of a project evaluating the effectiveness of a new statistics curriculum in US high schools (Konold, 1989; Garfield, 1996, 1998a, 2003). Each item in the SRA describes a statistics or probability problem and offers four to eight choices of responses. Most responses include a statement of reasoning, explaining the rationale for a particular choice. One response of every item corresponds to a category of correct reasoning; all or most other responses correspond to categories of misconceptions. For a full description of the individual items and the eight correct reasoning scales and eight misconceptions scales, we refer to Garfield (1998a, 2003); Table 3.1 summarizes the several scales of the instrument. In the design process of the instrument, the authors included several stages directed at achieving good validity and reliability: see Garfield (2003). With regard to criterion-related validity, Garfield (2003) reports extreme low correlations with different course outcomes, suggesting statistical reasoning and

misconceptions to be unrelated to course performance. In addition, Garfield (2003) reports satisfactory test-retest reliabilities, but low internal consistency reliability coefficients, implying that reasoning scales and misconception scales respectively appear not to measure one single ability or trait.

In terms of the classification into the more recently developed categories of statistical literacy, reasoning, and thinking, the allocation of individual reasoning abilities and misconceptions to these three classes is not obvious. Aver, TWay, AverMc, High%, and Groups refer to basic data-related skills, and seem to fit best in the literacy category. At the other extreme, Comp, Sampl, Correl, Small, Cause, and EquiPr involve probability and statistical theory related concepts, and might better suite the thinking category. The remaining scales, referring to notions of probability and uncertainty, would then fit the reasoning category. We will return to this issue when discussing descriptive statistics of SRA data obtained from this study and a limited number of other studies that provide empirical data on the instrument: Garfield (1998b, 2003), Garfield and Chance (2000), Liu (1998) and Sundre (2003).

Table 3-1. SRA Correct reasoning scales and misconceptions scales; based on Garfield (2003)

Correct Reasoning Scales:

Prob: Correctly interprets probabilities. Assesses the understanding and use of ideas of randomness, chance to make judgments about uncertain events.

Aver: Understands how to select an appropriate average. Assesses the understanding what measures of center tell about a data set, and which are best to use under different conditions.

Comp: Correctly computes probability, both understanding probabilities as ratios, and using combinatorial reasoning. Assesses the knowledge that in uncertain events not all outcomes are equally likely, and how to determine the likelihood of different events using an appropriate method.

Indep: Understands independence.

Sampl: Understands sampling variability

Correl: Distinguishes between correlation and causation. Assesses the knowledge that a strong correlation between two variables does not mean that one causes the other.

2Way: Correctly interprets two-way tables. Assesses the knowledge how to judge and interpret a relationship between two variables, knowing how to examine and interpret a two way table.

LrgS: Understands the importance of large samples. Assesses the knowledge of how samples are related to a population and what may be inferred from a sample; knowing that a larger, well chosen sample will more accurately represent a population; being cautious when making inferences made on small samples.

Misconception scales:

AverMc: Misconceptions involving averages. This category includes the following pitfalls: averages are the most common number; failing to take outliers into consideration when computing the mean; comparing groups on their averages only; and confusing mean with median.

OutcO: Outcome orientation. Students use an intuitive model of probability that lead them to make yes or no decisions about single events rather than looking at the series of events; see Konold (1989).

High%: Good samples have to represent a high percentage of the population. Size of the sample and how it is chosen is not important, but it must represent a large part of the population to be a good sample.

Small: Law of small numbers. Small samples best resemble the populations from which they are sampled, so are to be preferred over larger samples.

Repre: Representativeness misconception. In this misconception the likelihood of a sample is estimated on the basis how closely it resembles the population. Documented in Kahneman, Slovic, & Tversky (1982).

Cause: Correlation implies causation.

EquiPr: Equiprobability bias. Events of unequal chance tend to be viewed as equally likely; see Lecoutre (1992).

Groups: Groups can only be compared if they have the same size.

Attitudes and beliefs toward statistics. Attitudes are measured with the Survey of Attitudes Toward Statistics (SATS) developed by Schau and co-authors (Schau, Stevens, Dauphinee, & DelVecchio, 1995; Dauphinee, Schau, & Stevens, 1997; Sorge & Schau, 2002; Schau, 2003; Hilton, Schau, & Olsen, 2004). There are two existing versions of the SATS, both consisting of seven-point Likert-type items measuring aspects of post-secondary students' statistics attitudes. The 28-item version of SATS contains four scales, as indicated below. Each scale is accompanied by two examples of items, one positively and one negatively worded:

- Affect (six items): measuring positive and negative feeling concerning statistics, the enjoyment aspect of intrinsic value: I like statistics; I am scared by statistics.
- Cognitive Competence (six items): measuring attitudes about intellectual knowledge and skills when applied to statistics, the self-concept of one's ability component in the expectancy-value model: I can learn statistics; I have no idea of what's going on in statistics.
- Value (nine items): measuring attitudes about the usefulness, relevance, and worth of statistics in personal and professional life, the utility and attainment components of task value: I use statistics in my everyday life; I will have no application for statistics in my profession.
- Difficulty (seven items): measuring attitudes about the difficulty of statistics as subject, the perception of the task demand: Statistics formulas are easy to understand; Statistics is highly technical.

Schau et al. (1995), Dauphinee et al. (1997), and Harris and Schau (1999) elaborate on the development process of the instrument. The instrument is freely available from the internet (Schau et al., 1999). Validation research has shown that a four-factor structure provides a good description of responses to the SATS-instrument in two very large samples of undergraduate students (Dauphinee et al. 1997, Hilton et al., 2004).

Recently, Schau has developed a 36-item version of the SATS, containing two additional scales, each covered by four, positively worded, items (Schau, personal communication, November 30, 2003). These scales, with one item example, are:

- Interest (four items): students' level of individual interest in statistics, the interest aspect of intrinsic value: I am interested in learning statistics.

- Effort (four items): amount of work the student expends to learn statistics, the perceived cost component of task value: I plan to work hard in my statistics course.

2.3 Data analysis

Parceling. The very first step in the data analysis is to reverse of negatively worded items in the SATS instrument, such that for all items a higher score corresponds to a more positive attitude. This step is worthwhile to mention since it requires attentiveness in the interpretation of the construct Difficulty. High scores for Difficulty express a more positive attitude, implying that a better name for the Difficulty scale would have been 'perceived lack of difficulty'. The second step of analysis is the parceling of the SATS data, following earlier empirical work by Schau and co-authors (Schau et al., 1995; Dauphinee et al., 1997; Hilton et al., 2004). The technique of item parceling, where items from the same subscale are aggregated into several parcels or miniscales, has been adopted in empirical studies for several reasons: obtaining more continuous and normally distributed observed data, reducing the number of model parameters to achieve a more attractive variable to sample size ratio, and to get more stable parameter estimates (Bandalos, 2002; Hau & Marsh, 2004; Marsh et al., 1998).

In parceling items, Hau and Marsh (2004) advise not to reduce the number of indicators for each latent construct beyond a minimum of three. Next, they recommend to counterbalance skewness in case of strong non-normality by creating parcels out of item pairs with opposite skew. In order to determine the relevance of this recommendation of counterbalancing skewness for our data set, the degree of non-normality of the data was calculated as a preliminary step to parceling. In the data of the first four SATS factors, no indications of non-normality were found in any of the self-report questionnaires beyond Hau and Marsh's (2004) category of 'moderately non-normal', implying skew = 1.0 and kurtosis = 1.5. Items corresponding to the constructs Interest and especially Effort are however much stronger skewed.

In the empirical analyses of their 28-item SATS data, Schau et al. (1995), Dauphinee et al. (1997), and Hilton et al. (2004) adopt an item parceling scheme based on balancing with respect to the positively and negatively worded items, size of parcel means, standard deviations, and skew (see Schau et al., 1995). Their parceling solution contains two parcels for Affect, Cognitive Competence, and Difficulty each; only Value contains three. Given the rule of thumb of at least three parcels per factor and the

advice to counterbalance skew as much as possible, it was decided to apply a parceling scheme different from Schau and co-authors, based only on skewness, and resulting in exactly three parcels per factor.

Statistical analyses. This study integrates several techniques of structural equation modeling (SEM). A SEM model is distinct from a path or regression model in that it hypothesizes that crucial variables, such as attitudes in this study, are not directly observable and are better modeled as latent variables than as observable ones. In doing so, a SEM makes it possible to distinguish two different types of errors: errors in equations, as does the path model, and errors in the observation of variables. Making this distinction is especially worthwhile, when errors in important constructs have rather different sizes. Studying reliabilities of several achievement motivations, and their variation over subjects, suggests that this argument applies to this study. In our study, SEM's were estimated with LISREL (version 8.54) using maximum likelihood estimation. For further discussion of SEM, see e.g. Byrne (1998), Kline (2005), and Schumacker and Lomax (2004).

The standard approach to estimate a SEM distinguishes two steps (Schumacker and Lomax, 2004). In the first phase of the two-step model building approach, measurement models for all latent variables in the model are estimated. Measurement models are in general factor models that allow factors, also called traits, and sometimes uniqueness, i.e. the errors in indicators, to be correlated. In our study, we need to estimate three of such 'correlated trait' (CT), correlated uniqueness (CU), 'confirmatory factor analysis' (CFA) models: for the SATS data, for the SRA data, and for course performance data. In the second model building step, the structural part of the SEM is estimated. This structural part specifies the relationships between the independent and dependent latent variables. In contrast to the estimation of the measurement models, the estimation of structural relationships is to some extent explorative in nature. The structural part of the full structural equation model is not a priori restricted, except for several hypotheses with regard to the direction of the relationship. For the estimation of these structural parts, two different model modification procedures are applied. The first is called model trimming (Kline, 2005) or backward search (Schumacker & Lomax, 2004). Starting from a full matrix of structural path coefficients, one by one, parameters are restricted to zero if they prove non-significant, until all remaining structural parameters are significant. The second approach is called model building (Kline, 2005) or

forward search (Schumacker & Lomax, 2004). It starts from a zero matrix of structural paths coefficients, and frees parameters one by one, in the order indicated by the value of the modification indices, up to point where no more significant improvement in fit is achieved. Since in both approaches subsequent models are nested, the chi-square difference statistic can be used to assess model fit. In all five subjects, both forward and backward searches converge to the same final model. Model modification is a form of explorative analysis, and brings along the risk of capitalization on chance.

With large sample sizes as in our study, the χ^2 test statistic is known to always reject in any formal test of significance (Byrne, 1998; Marsh & Yeung, 1996). For that reason, and following Marsh & Yeung (1996), and Hilton et al. (2004), emphasis is placed on the Root Mean Square Error of Approximation (RMSEA), the Goodness-of-Fit Index (GFI), the Non-Normed Fit Index (NNFI; termed Tucker-Lewis Index or TLI in Marsh & Yeung, 1996), the Comparative Fit Index (CFI) and the Relative Fit Index (RFI, termed Relative Noncentrality Index or RNI in Marsh & Yeung, 1996), and the normed version of the χ^2 test statistic: χ^2/df . For the last index, no clear-cut guidelines exist: values in the range of 2.0 to 5.0 are acceptable, with lower values indicating better fit. For RMSEA, values ≤ 0.05 indicate good fit, values ≤ 0.08 indicate reasonable fit. The indices GFI, NNFI, CFI, and RFI, all normally lie in the range 0.0 - 1.0, with higher values indicating better fit. As a benchmark for good fit, the value 0.90 is often used (Kline, 2005).

3. RESULTS

3.1 Descriptive statistics of attitudes and beliefs toward statistics

Descriptive statistics of the SATS scales are exhibited in Table 3.2 and Figure 3.1. All attitudes are measured using a Likert 1-7 scale. Since all scale means, except for Difficulty, are larger than the neutral value of four of the used Likert scale, students in our sample express positive attitudes toward statistics for Affect, Cognitive Competence, Value, Interest, and Effort. Means and standard deviations are in line with values reported in Schau (2003) found as pre-test scores in a large class of undergraduate US students: Affect, Cognitive Competence, and Value are slightly more positive in our sample, Difficulty is equal. In comparing our European data

with data from US studies, it is important to realize that participants in our study are all in economics and business programs. These programs require students to take math classes in high school on at least intermediate level. Cronbach α reliability coefficients of these four scales are satisfactory, and again in line with intervals of values reported in Schau (2003) from several empirical studies by Schau and co-researchers. No empirical studies exist at this moment that incorporate the new scales of the 36-item SATS version: Interest and Effort. In our study, both these attitudes are clearly positive on average, with (planned) Effort taking a very strong position with a mean of 6.37 on a 1-7 scale. Figure 3.1 indicates that due to the high scores on Effort, skewness is an issue for this scale, and not for the other scales.

Table 3-2. Scale means, standard deviations, and Cronbach α 's for attitudes toward statistics in our study (N=1458) and as reference, values reported in Schau (2003)

	Mean (Standard deviation)		Cronbach α	
	this study	Schau (2003)	this study	Schau (2003)
Affect	4.52 (1.10)	4.03 (1.14)	0.82	0.80 - 0.89
Cognitive Competence	5.08 (0.89)	4.91 (1.09)	0.78	0.77 - 0.88
Value	5.05 (0.83)	4.86 (1.01)	0.78	0.74 - 0.90
Difficulty	3.59 (0.77)	3.62 (0.78)	0.68	0.64 - 0.81
Interest	5.07 (0.99)		0.80	
Effort	6.37 (0.72)		0.76	

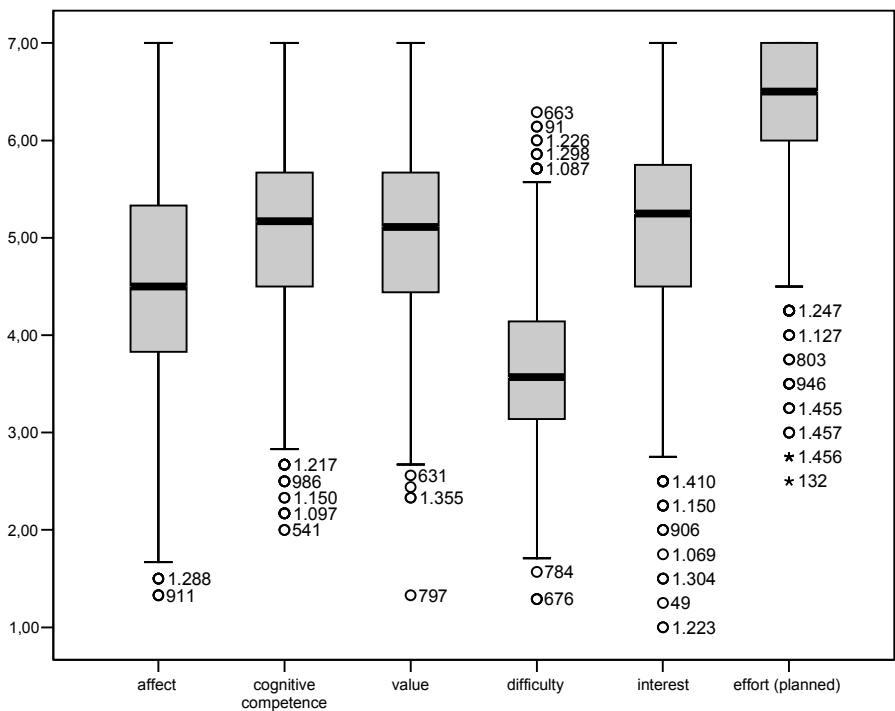


Figure 3-1. Descriptives of SATS scales (N=1458)

3.2 Descriptive statistics of statistical reasoning abilities

Descriptive statistics of the SRA data, similar to those reported in Garfield (1998b, 2003), Garfield and Chance (2000) and Liu (1998), are exhibited in Table 3.3. Since the maximum score of the several scales varies with the total number of answer options corresponding to the scale, the table presents the means of the several scales expressed as a proportion, that is on a [0-1] scale. In addition to scores on eight reasoning skills, and eight misconceptions, the aggregated correct reasoning score (Correct) and aggregated misconceptions (Misconcep) are reported. The aggregated scores are obtained in the same way as in the studies by Garfield and co-authors by taking the sum over all correct reasoning and misconception items, and re-expressing them as a proportion. Since the number of items per scale ranges from 1 to 5, different scales have a different weight in the

total score, so aggregated scores are to be regarded as weighted averages. Data reported by Garfield and co-authors are restricted to means.

Table 3-3. Scale means and standard deviations for statistical reasoning abilities in our study (n=1499) and as reference, post-course values US college students reported in Garfield (2003)

	Mean (Standard deviation)			Mean (Standard deviation)	
	this study	Garfield (2003)		this study	Garfield (2003)
CC1	0.75 (0.29)	0.68	MC1	0.46 (0.27)	0.30
CC2	0.71 (0.27)	0.61	MC2	0.22 (0.17)	0.23
CC3	0.40 (0.25)	0.46	MC3	0.15 (0.23)	0.09
CC4	0.64 (0.29)	0.63	MC4	0.28 (0.27)	0.29
CC5	0.28 (0.30)	0.22	MC5	0.12 (0.22)	0.17
CC6	0.66 (0.47)	0.51	MC6	0.28 (0.37)	0.10
CC7	0.74 (0.40)	0.65	MC7	0.57 (0.33)	0.56
CC8	0.71 (0.33)	0.68	MC8	0.29 (0.46)	0.60
CC	0.58 (0.13)	0.55	MC	0.29 (0.10)	0.27

Outcomes of our study and those reported in Garfield (2003) are remarkably similar, although the composition of groups of participating students is rather different. Garfield’s study refers to US college students surveyed at the end of an introductory course statistics, our study to European university students at the start of such an introductory course. Of the correct reasoning scales, Prob and Twow are amongst those with highest mastery level, and Comp and Sampl with lowest. Of the misconception scales, EquiPr and Groups are high in all studies (in our sample, Groups somewhat less), and High%, Repre and Cause are low.

Conceptions for which we find higher scores than reported in the Garfield-studies are Aver, Correl, and Twow. The misconception for which our data indicates a remarkable low relative score is Groups. Of these four scales, three are characterized earlier as being part of the category of statistical literacy. This meshes the difference in timing of the instrument, as a pre-test in our study, and a post-test in other studies. Not (recently) educated in introductory statistics, it is not surprising that students in our study score relative high on statistical literacy components, but low on a statistical thinking related component as MC6, correlation implies causation, typically an important concept to be taught in an introductory course.

As a last observation on average levels of reasoning skills and misconceptions, the high rate of correct answers is noticeable. Of the eight correct reasoning skills, five have means of above 65% correct. Of the eight misconception scales, only two have means larger than 30%.

3.3 Measurement model of attitudes and beliefs toward statistics

As a first step in the modeling of SATS data, an explorative factor analysis was performed (principal components, varimax rotation). The eigenvalue-criterium identifies six factors. The scree-criterium demonstrates a large jump at four factors, and a smaller jump at six factors. The newly created scales Interest and Effort clearly qualify as independent factors. The same is true for the scale Value. However, items in the scales Affect, Cognitive Competence, and Difficulty are strongly correlated. This finding coincides with other empirical studies on SATS: Schau et al. (1995), Dauphinee et al. (1997), Hilton et al. (2004), and Cashin and Elmore (2005). On the basis of these high correlations, Cashin and Elmore (2005) decide to reduce the three scales Affect, Cognitive Competence, and Difficulty into one latent factor, whereas in the other three studies they are modeled as separate, but correlated, latent factors. We follow the last approach: a six-factor confirmatory factor model was estimated on parcelled attitudes data allowing a correlated traits (CT) structure but no cross-loadings in the factor loading matrix and no correlated uniqueness (CU) factor. Table 3.4 contains fit indices of this CT factor model, Figure 3.2 the structure of the factor model, including estimated trait correlations.

Table 3-4. . Fit indices of six-factor correlated traits confirmatory factor models of attitudes toward statistics

	χ^2	df	RMSEA	GFI	NNFI	CFI	RFI
CT 6CFA model	701.80	123	.057	.95	.97	.97	.96

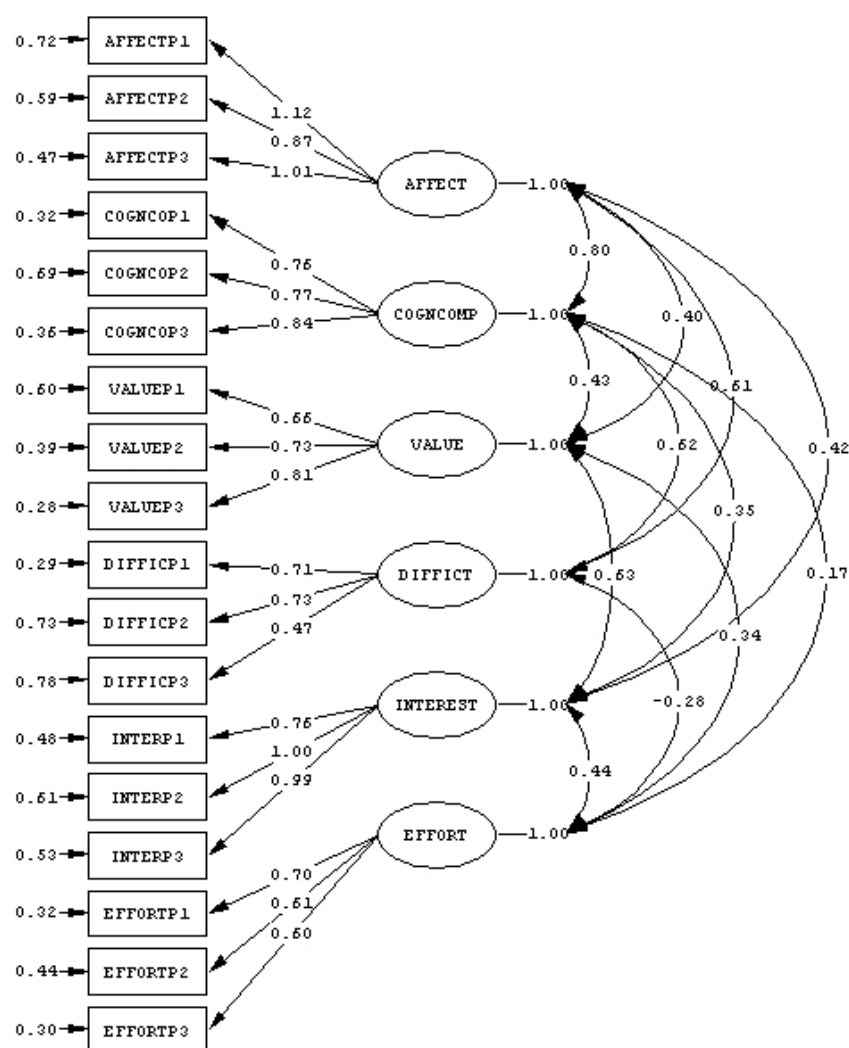


Figure 3-2. Correlated traits factor model as measurement model for attitudes toward statistics. Values are standardized parameter estimates. All values shown are statistically significant, $p < .05$. AFFECT, Affect; COGNCOMP, Cognitive Competence; VALUE, Value; DIFFICT, Difficulty. INTEREST, interest; EFFORT, (planned) effort

Fit indices indicate that the hypothesized correlated traits factor model fits the data quite well. Having confirmed the six-factor model, the correlation

structure of latent factors depicted in Table 3.5 deserves prime interest. Table 3.5 demonstrates that twelve out of fifteen trait correlations are significant. Only three trait correlations appear to be non-significant and are restricted to zero in the estimation of the final version of the factor model, with the other correlations freed.

Table 3-5. Estimated latent factor correlations of attitudes toward statistics

	Affect	Cognitive Competence	Value	Difficulty	Interest	Effort
Affect	1.00					
Cognitive Competence	.80	1.00				
Value	.40	.43	1.00			
Difficulty	.61	.62	-	1.00		
Interest	.42	.35	.63	-	1.00	
Effort	-	.17	.34	-.28	.44	1.00

Note: all correlations are significant at $p < .000001$.

When interpreting the trait correlation structure, the first issue that comes up is the effect of disentangling the broad task value concept into, Affect, related to liking the subject, and Value, related to the importance attached to the subject. The correlation between latent factors Affect and Value ($r = .40$) is, relative to other correlations, modest: .40. This indicates that Affect and Value are clearly empirically distinguishable constructs. The correlation between Value and Difficulty is insignificant, indicating that the attached value is independent to the lack of perceived difficulty. A third observation refers to, the by far largest, correlation; namely between Affect and Cognitive Competence. This is in itself a remarkable fact: Affect is achieved by decomposing the task value component into affective and utility-related factors, but from this analysis it appears that Affect is much stronger related to the expectancy component Cognitive Competence, than to Value. This once again confirms the usefulness of the affect extension of the expectancy-value model. The strong correlation we find, is comparable to the results found in Dauphinee et al. (1997), and Hilton et al. (2004).

The relationship between the two factors Interest and Effort and the other four factors is primarily through Value. Interest is unrelated to Difficulty, and Effort is unrelated to Affect and negatively related to Difficulty. That last negative relationship seems to be an exponent of rational study behavior: students who regard statistics as difficult, plan to invest more study efforts than students regarding the subject as less difficult. However, it is at odds with the expectancy-value model, where that

relation has the opposite sign. The different outcome is best explained by the context in which the model is used; whereas the expectancy-value model is primarily based on the selection of learning tasks (such as choosing one course in favor of another), the context of this study is the intensity of performance, given the required learning tasks. In the expectancy-value model, Effort is assumed to be an intermediate outcome variable. For this interpretation to be true, the correlations between Effort and its predictors are expected to be strongly positive. This is not the case, except possibly for Interest. Two potential explanations for the weaker than expected relationship between Effort and its predictors are available. First, Effort is an ex-ante measure, and planned effort might quite well diverge strongly from ex-post measured, realized effort. Second, planned Effort scores seem to be a composition of two rather different underlying mechanisms that can make the relationships of this variable to other attitudinal constructs ambiguous. On the one side, students with high achievement motivation are assumed to spend high efforts in their learning, so planned effort acts as a proxy for achievement motivation. On the other side, planned effort might act as a proxy for students' learning approaches: students with a tendency to a memorizing type of learning tend to invest more efforts in their learning than students with a learning approach focused on understanding. In general, the latter deep learning approach is regarded as better, and at least more efficient, than the first mentioned surface learning approach. For that reason, it might be expected that students with a tendency towards deep learning will have more positive attitudes, making deep learning positively related to the several attitudinal variables, and surface learning negatively related. If this is true, the relationship between Effort and attitudinal variables is the resultant of two counterbalancing forces: higher planned effort levels when being motivated, but lower planned effort levels when relying on efficient, deep learning approaches. In the subsection discussing the outcomes of the full structural equation model, we will further elaborate this issue.

3.4 Measurement model of statistical reasoning abilities

Previous empirical studies of the SRA instrument have used aggregated correct conceptions, and aggregated misconceptions, as scales, with the eight correct reasoning ability scores and the eight misconception scores as items. This would suggest a measurement model with the two aggregated reasoning abilities as latent constructs, and the correct reasoning ability

and aggregated variables as indicators. However, Garfield (1998b), Garfield and Chance (2000) and Liu (1998) point out that this modeling approach has important drawbacks: in their studies, as in ours, the correlations between reasoning ability scores are low, mostly insignificant, and quite often of opposite signs. This is problematic in terms of scale construction, since it gives rise to low values of instrument reliability. In the present data set analyzed in this study, the Cronbach- α reliabilities of the correct reasoning scales is 0.34, while for the misconception scales, the reliability α is 0.10. These values are too low to warrant meaningfulness of aggregated constructs. Elsewhere, we have investigated the reliability of aggregated scales for a much larger sample, and came to similar conclusions (Tempelaar, 2004). Deleting individual items with extreme p -values, as suggested in Liu (1998), appears to have little impact on reliabilities in our data.

Inspection of the correlation matrix depicted in Table 3.6 does however expose a pattern in correlations that suggests an alternative approach to model the outcomes of the SRA-instrument. Correlations within the group of correct reasoning scales, and within the group of misconceptions are, without exception, low. However, in the rectangular part of the correlation matrix containing the correlations between correct reasoning skills and misconceptions, seven out of eight columns contain exactly one highly significant and strongly negative correlation. This is not surprising: from the definition of e.g. Prob and OutcO it is apparent that outcome orientation, that is the use of an intuitive and incorrect probability model, is at odds with correctly interpreting probabilities. And in some cases, the strong negative correlations between several correct conceptions and misconceptions find their origin in the fact that the concepts are based on different options of the same multiple choice items, which would lead to negative correlations by construct (although several multiple choice items allow for multiple answers).

Table 3-6. Correlations between SRA correct reasoning (CC) and misconceptions (MC) scales being significant at $p = 0.01$; values in bold exceed 0.30 in absolute value

	Prob	Aver	Comp	Indep	Sampl	Correl	Twow	LrgS
Prob	1.00							
Aver		1.00						
Comp	.09		1.00					
Indep		.08	-.16	1.00				
Sampl		.10	.08	-.07	1.00			
Correl	.09	.17				1.00		
Twow	.13	.13				.09	1.00	
LrgS		.10	.09		.07	.09	.09	1.00
AverMc		-.43			-.26			
OutcO	-.42		-.22	-.13				-.32
High%						.08		.11
Small		-.10	-.09	.07	-.69			-.16
Repre			-.21	-.69	.08			
Cause		-.08			-.07	-.46		
EquiPr			-.80	.20	-.12	.09		
Groups								

	AverMc	OutcO	High%	Small	Repre	Cause	EquiPr	Groups
AverMc	1.00							
OutcO		1.00						
High%			1.00					
Small				1.00				
Repre					1.00			
Cause	.14		-.07		.10	1.00		
EquiPr				.12	-.10		1.00	
Groups	.07		.09					1.00

Taking this pattern of correlations into account, a different method of aggregating scales scores than calculating total correct and misconception scores suggests itself. On the basis of the strong negative correlations between seven pairs of one correct reasoning scale and one misconception scale, a pair-wise aggregation process seems to be more appropriate than aggregation over all correct, and all incorrect answers. To investigate this option, an exploratory factor analysis is performed. This factor analysis results in a seven-factor solution, with five factors composed by pairs of one correct conception and one misconception, having factor loadings of opposite signs: Comp & EquiPr, Sampl & Small, Indep & Repre, Prob &

OutcO, and Correl & Cause. The remaining two factors are composed of Aver, Twow, LrgS, & AverMc; and High% & Groups, respectively. All factor loadings have expected signs: positive for correct conceptions, negative for misconceptions.

Subsequently, a measurement model is estimated taking the outcome of the explorative factor analysis as its basis. No cross-loadings were allowed but, similar to the estimation of the attitudes measurement model, trait correlations are allowed. In addition, uniqueness correlations are allowed for those reasoning abilities and misconceptions that share an item. Of the 21 trait correlations, only four appear to be significant. This does not come as a surprise, given the many insignificant correlations in Table 3.6. All 10 uniqueness correlations appear to be significant. The final measurement model for reasoning abilities is depicted in Figure 3.3; the fit indices of the final model are reported in Table 3.7. The fit of the CTCU 7 CFA model is good.

Table 3-7. Fit indices of seven-factor correlated traits confirmatory factor models of statistical reasoning abilities

	χ^2	df	RMSEA	GFI	NNFI	CFI	RFI
CTCU 7CFA model	355.00	98	.042	.97	.93	.94	.90

Judging from the good fit of this measurement model, an important conclusion with regard to the SRA instrument becomes apparent. When using SRA as an instrument to assess statistical reasoning, it is less attractive to aggregate all correct scales and all misconception scales into constructs like total correct reasoning and total misconceptions, given the limited reliability of such constructs. As an alternative, composing latent reasoning constructs on which both correct and misconception scales load seems to offer higher reliability.

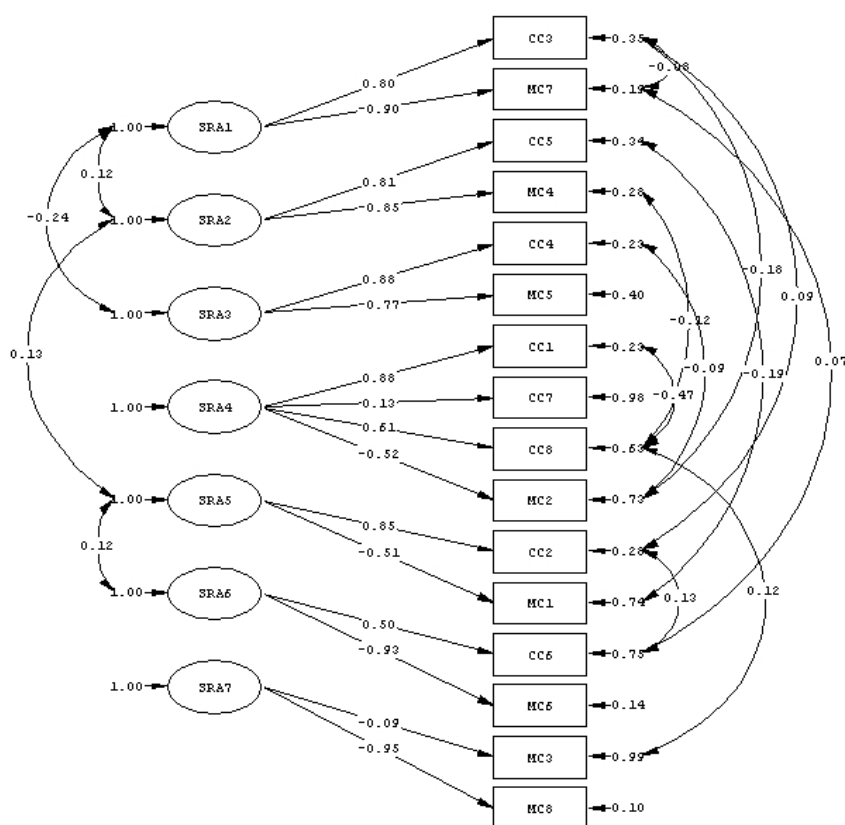


Figure 3-3. Correlated traits, correlated uniqueness factor model as measurement model for statistical reasoning abilities. Values are standardized parameter estimates. All values shown are statistically significant, $p < .05$. CC, Correct Conception; MC, MisConception; SRA, latent reasoning factor

3.5 Full structural equation model of attitudes and beliefs toward statistics, statistical reasoning abilities, and course performance

The final step in the analysis regards the integration of both measurement models. This includes the not explicitly elaborated model for course performances, specifying the two latent course performances EXAM and QUIZ. Both course performance constructs are measured by two indicators: a score for mathematics, and a score for statistics. The relationships that

link the latent factors in the three measurement parts constitute the structural part of the model. The estimation of the structural parameters is similar to the estimation of trait correlations in the measurement models; no a priori restrictions apply as to what parameters are restricted to zero and which are set free. Two modification directions were applied: model building and model trimming. Both methods converge to the model depicted in Figure 3.4. Figure 3.4 does not make explicit the estimated correlations between latent factors; the same correlation structure as visible in Figures 3.2 and 3.3 was however used in the estimation of the full model. Table 3.8 reports fit indices of that model and indicates good fit. Table 3.9 describes the standardized parameter estimates or β -coefficients of the structural part of the model.

Table 3-8. Fit indices of full structural model of attitudes toward statistics, statistical reasoning abilities, and course performance

	χ^2	<i>df</i>	RMSEA	GFI	NNFI	CFI	RFI
SEM	1599.26	620	.035	.94	.96	.97	.94

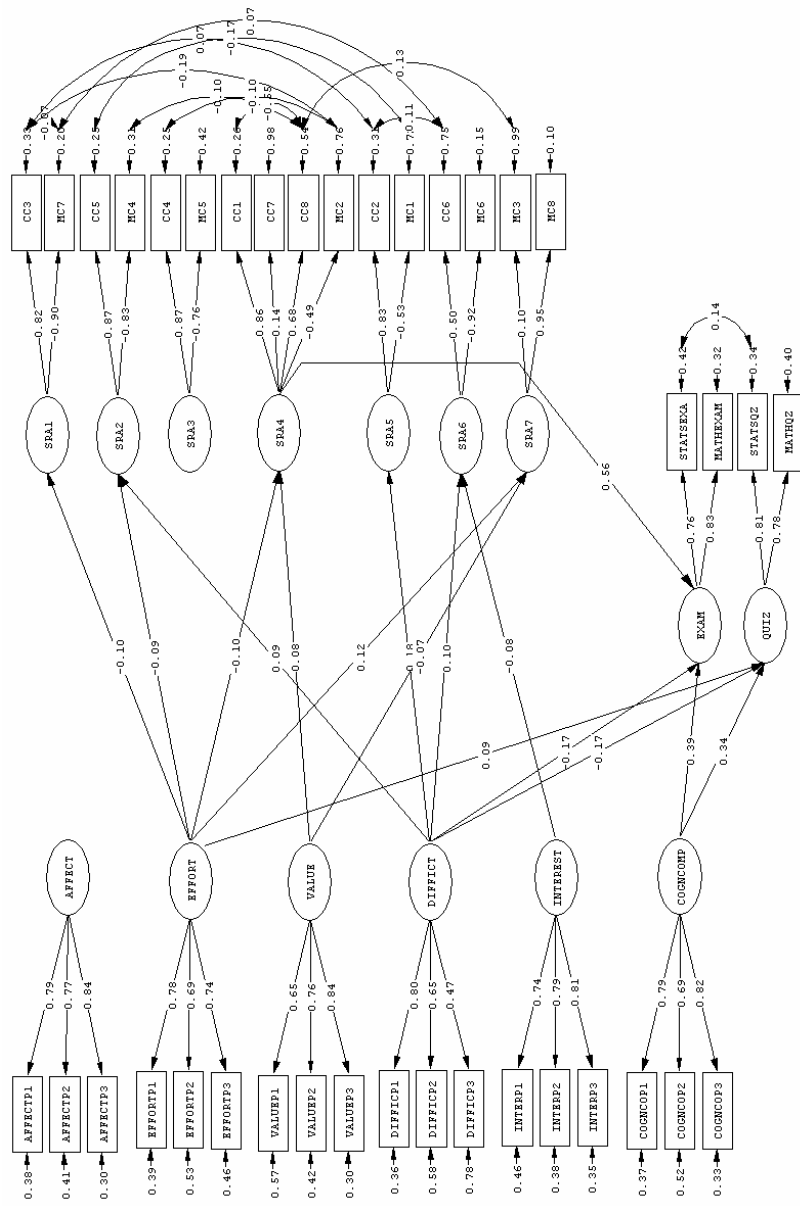


Figure 3-4. Structural equation model of attitudes toward statistics, statistical reasoning abilities, and course performance. Values are standardized parameter estimates. All values shown are statistically significant, $p < .05$. AFFECT, Affect; COGNCOMP, Cognitive Competence; VALUE, Value; DIFFIC, Difficulty. INTEREST, interest; EFFORT, (planned) effort, SPA1..7, latent reasoning factors; EXAM, QUIZ, latent course performance factors

Table 3-9. Standardized estimates of the structural part of the full structural model of attitudes toward statistics, statistical reasoning abilities, and course performance

	Affect	Cognitive Competence	Value	Difficulty	Interest	Effort	SRA4
SRA1						-0.10	
SRA2				0.09		-0.09	
SRA3							
SRA4			0.08			-0.10	
SRA5				0.18			
SRA6				0.10	-0.08		
SRA7			0.07			-0.12	
EXAM		0.39		-0.17			0.56
QUIZ		0.34		-0.17		0.09	

The final structural model allows several interpretations. Students’ self-ability belief Cognitive Competence is a strong predictor of both latent course performance factors, with β -values of 0.39 and 0.34. This is in agreement with many studies on the expectancy-value model, and self-concept or self-efficacy research. The relationships between statistical reasoning and the two course performance factors are weak, what is in line with the low correlations between SRA constructs and course performance found in several studies. In our study, only SRA4, the latent factor composed of four correct conceptions and misconceptions related to the ability to interpret probabilities, has a significant and strong impact on the latent exam factor.

The second direct effect from attitudinal variables on course performances stems from the other expectancy construct of perceived task demand: (lack of) Difficulty. The relationship is reversed, with β -values of -0.17. This outcome is somewhat surprising: the expectancy-value model would predict a positive relationship. However, the relationship is robust: using a split-sample approach (and path analysis), it is confirmed in subsamples composed in several ways. The bivariate relationship between Difficulty and performance is however absent; the negative relation we find is only present in a simultaneous relation between Cognitive Competence, Difficulty, and course performance. It should thus be interpreted as a process of underestimation of task demand by students with an above average ability belief.

The reduced form squared multiple correlations of both course performance latent factors EXAM and QUIZ are equal to 0.10. This means

that the combined effect of both direct paths from SATS variables to EXAM and QUIZ, and the indirect paths from SATS via SRA to the two course performance factors, explains 10% of total variation in both course performances. In the decomposition of explained variation into direct and indirect effects, it becomes clear that the contribution of the indirect effect is neglectable: less than 0.5%. The dominance of direct over indirect effects is due to the fact that relations between SATS and SRA are weak, and much weaker than relations between SATS and performance. In line with the expectancy-value model, attitudes have a positive impact on reasoning abilities through the variables Value and (perceived lack of) Difficulty. In contrast to predictions based on the expectancy-value model, the Effort variable has a negative impact on four of the seven latent reasoning factors. The negative relationship is consistent: β -coefficients of Effort to the several SRA's are either significantly negative, or zero, but never positive. Although a negative relation may appear counter-intuitive, it is in line with related research on the relationship between preferred learning approaches and reasoning abilities, where it was found that a tendency to surface learning negatively influences statistical reasoning (Tempelaar, 2004; Tempelaar et al., 2006; Tempelaar, Schim van der Loeff, Gijsselaers, Crombrugghe, 2007). Planned effort being a proxy of both achievement motivation and a non-efficient learning approach, see the above discussion of the measurement model of attitudes, will give rise to diverse relationships between learning outcomes and the Effort variable. Learning performances that allow for alternative learning paths - such as memorizing versus understanding - are expected to demonstrate a positive relationship with planned effort. For these learning performances, the achievement motivation component in planned effort is dominant: students, who are prepared to work hard, will achieve better performances. In our study, quiz scores for both mathematics and statistics are the ultimate example of such type of course performances. Quizzes are designed to be accessible for all students and the bonus points they bring about are especially helpful for students at risk in passing the course. This makes plausible that the motivation component in planned effort dominates the learning approach component, which explains the positive relationship between Effort and Quiz. The opposite case is constituted by the SRA factors. Since the SRA is administered as an entry measurement unrelated to course grading, any direct effect of achievement motivation can be assumed to be absent. And since statistical reasoning is not part of any secondary education of most students in this study, indirect effects - taking advantage of having been

highly motivated in secondary school - will at most be very modest. As a result, the learning approach component in planned effort is expected to be dominant, what quite well explains the negative relationships found between EFFORT and four of the SRA factors. In this spectrum of course performances, the scores in the exam take an intermediate position. Being the course performance measurement, they certainly contain a strong achievement motivation component. At the same time, exams are certainly much less accessible than quizzes, what feeds the learning approach component. In the aggregation, the two effects are counterbalancing, what quite well might explain the latent factor EXAM being unrelated to Effort.

4. CONCLUSIONS

In this study the affect-extended version of the expectancy-value model (Schau et al., 1995; Dauphinee et al., 1997; Hilton et al., 2004) was adopted as achievement motivation model. Our data corroborate this extension, in the sense that affect and value turn out to be clearly distinguishable constructs, as well as in the sense that these variables play a distinctive role in the relationships with reasoning abilities and course performance. To our knowledge this study is the first to apply the 36-item SATS version, with the new scales Interest and Effort. Both scales appear to be a valuable addition to the instrument. The latent trait correlations in Table 3.5 demonstrate that the two factors are well identified constructs. However, correlational analysis suggests that Effort might be composed of two rather different characteristics. Therefore, a decomposition of this scale into an achievement motivation aspect, and a learning approach aspect, is called for. The latter aspect has the interpretation that students with a surface learning approach will typically achieve high scores on this Effort variable, since they invest large amounts of time for learning subjects by memorization.

Through a factor analytic study, we conclude that a factor model with most factors being composed of pairs of one reasoning ability and one misconception, provides an appropriate measurement model. This shows that the SRA-instrument used by Garfield (1998b, 2003), Garfield and Chance (2000) and Liu (1998) is not flawed. In studies by these authors only two aggregate scales, one for statistical reasoning abilities, and one for statistical misconceptions, are employed. They point out that these aggregate scales have shortcomings in view of the low values of correlations between the scales that constitute both aggregate scales,

which results in low reliabilities. Our results imply that the finding of low correlations does not invalidate the instrument, but that alternative measurement models other than the one based on aggregate scales should be used.

This study adds support to previous findings of the absence of a strong relationship of misconceptions and their counterpart, the reasoning abilities, with students' course performances. This is demonstrated in studies where statistical reasoning is regarded as one of the several learning outcomes of the course and assessed simultaneously with these other course performances (Garfield, 1998b, 2003; Garfield and Chance, 2000; and Liu, 1998). In this study, Tempelaar (2004) and Tempelaar et al. (2006) it is also demonstrated in a second type of studies, where statistical reasoning is regarded as part of the prior knowledge state of the student and assessed before the start of the course. Are these studies, given their conclusions that SRA components are only weakly or even un-related to different course performance indicators, uninformative? We would argue that the opposite is true: exactly because of these absent relationships, they are informative. In general, different components of statistical knowledge, measured as course performance scores, tend to be substantially correlated. For example, in this study the correlation between latent course performance factors EXAM and QUIZ equals $r = 0.69$. And investigating the relationships between three rather different types of course performances, final exam scores, quiz scores, and home work scores, we find similar substantial correlations (Tempelaar et al., 2006). Since the SRA-instrument is developed to assess statistical reasoning mastery achieved in high school statistics programs, the natural hypothesis is that SRA-scores correlate with the several course performances in the same way as the other components of course performances do. But they clearly do not do so. It are these unexpected low correlations that make studies as ours informative, rather than the case that the expected, substantial positive correlations would have been found.

The absence of substantial relationships can be well explained in the context of naïve theories that is an element of the new theory of learning, as elaborated in Bruer's (1993) 'Schools for thought'. Naïve theories or misconceptions are informal, self-acquired elements of science knowledge, inconsistent with formal science. Students can possess formal knowledge and naïve knowledge at the same time: the learning of formal knowledge does not automatically imply that naïve knowledge is unlearned. In spite of having mastered the formal knowledge, students tend to solve scientific problems with their naïve knowledge, especially when they are confronted

with these problems outside a school context. And, worst of all, formal knowledge tends to get forgotten much faster than naïve knowledge. Empirical outcomes of studies using the SRA-instrument are in line with these observations. Absence of substantive relationships is compatible with the hypothesis that both statistical reasoning abilities and statistical misconceptions are part of students' naïve statistical knowledge: the first category naïve and correct, the second category naïve but incorrect. More research to investigate the role of naïve theories in learning and the development of naïve knowledge in time is necessary. This is particularly relevant since the reform movement in statistics education has called for a more prominent position of statistical reasoning, and the related domains of statistical literacy and thinking in the statistics curriculum. So it is the reformed curriculum, more than any traditional curriculum that requires resolving the instructional challenge of unlearning statistical misconceptions before being able to replace them by proper reasoning abilities.

Empirical studies as documented in *Statistics Education Research Journal* (2004:2; 2005:1) and Ben-Zvi & Garfield (2005) conclude that in order to learn reasoning and to unlearn misconceptions, the use of specific educational tools is indispensable. This study suggests that the use of these tools is probably only part of the solution of the instructional challenge. A strong dependency on these instructional tools might be at odds with educational principles on which student-centered programs are based, in the sense that they limit students' own responsibility to organize the learning process. The outcomes of this study might bring forward some further limitations. In most learning processes students enter the learning context with a given set of background characteristics, such as a preference for deep learning versus surface learning. Most of these contexts allow all students to achieve satisfactory learning outcomes, be it along different learning paths. As a concrete example: our structural equation model suggests that both surface learning oriented students and deep learning oriented students can achieve adequate course performance scores. But our empirical analyses also suggests that statistical reasoning might be the odd man out in this context: the learning of statistical reasoning seems not to easily assimilate to variation in students' background characteristics as preferred learning approach as is the case with other cognitive goals. If this conclusion is correct, it implies we need an even broader range of educational tools than already described in the sources referred above: above content, the tools should address general learning approaches.

REFERENCES

- Ben-Zvi, D., & Garfield, J. B. (Eds.). (2004a). *The challenge of developing statistical literacy, reasoning, and thinking*. Dordrecht: Kluwer Academic Publishing.
- Ben-Zvi, D., & Garfield, J. B. (2004b). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp 3-15). Dordrecht: Kluwer Academic Publishing.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Bruer, J. T. (1993). *Schools for thought: A science of learning in the classroom*. Cambridge, Mass.: The MIT Press.
- Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.
- Cashin, S. E., & Elmore, P. B. (2005). The Survey of Attitudes Toward Statistics scale: a construct validity study. *Educational and Psychological Measurement*, 65(3), 509-524.
- Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, 10 (3). [Online: <http://www.amstat.org/publications/jse/v10n3/chance.html>].
- Chance, B. L., & Garfield, J. B. (2002). New approaches to gathering data on student learning for research in statistics education. *Statistics Education Research Journal*, 1 (2), 38-41. [Online: [http://www.stat.auckland.ac.nz/~iase/serj/SERJ1\(2\).pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ1(2).pdf)].
- Dauphinee, T. L., Schau, C., & Stevens, J. J. (1997). Survey of Attitudes Toward Statistics: Factor structure and factorial invariance for women and men. *Structural Equation Modeling*, 4 (2), 129-141.
- DelMas, R. (2002). Statistical literacy, reasoning, and learning. *Journal of Statistics Education*, 10 (3). [Online: http://www.amstat.org/publications/jse/v10n3/delmas_intro.html & http://www.amstat.org/publications/jse/v10n3/delmas_discussion.html].
- DelMas, R. (2004a). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79-95). Dordrecht: Kluwer Academic Publishing.
- DelMas, R. (2004b). Overview of ARTIST website and Assessment Builder. *Proceedings of the ARTIST Roundtable Conference*, Lawrence University. [Online: <http://www.rossmanchance.com/artist/Proctoc.html>].
- Eccles, J. S. (2005). Subjective task value and the Eccles et al. model of achievement-related choices. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105-121). New York: The Guilford Press.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109-132.

- Gal, I. (2004). Statistical literacy, Meanings, components, responsibilities. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 47-78). Dordrecht: Kluwer Academic Publishing.
- Gal, I., & Garfield, J. B. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal & J. B. Garfield, *The assessment challenge in statistical education* (pp. 1-13). Voorburg: IOS Press.
- Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education*, 2 (2). [Online: <http://www.amstat.org/publications/jse/v2n2/gal.html>].
- Garfield, J. B. (1996). Assessing student learning in the context of evaluating a chance course. *Communications in statistics; Part A: Theory and methods*, 25, 2863-2873.
- Garfield, J. B. (1998). *Challenges in Assessing Statistical Reasoning*. AERA Annual Meeting presentation, San Diego.
- Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, 2 (1), 22-38. [Online: [http://www.stat.auckland.ac.nz/~iase/serj/SERJ2\(1\).pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ2(1).pdf)].
- Garfield, J. B., & Ahlgren, A. (1988). Difficulties in learning basic concepts in statistics: Implications for research. *Journal for Research in Mathematics Education*, 19, 44-63.
- Garfield, J. B., & Ben-Zvi, D. (2004). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, 4 (1), 92-99. [Online: [http://www.stat.auckland.ac.nz/~iase/serj/SERJ4\(1\)_Garfield_BenZvi.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ4(1)_Garfield_BenZvi.pdf)]
- Garfield, J. B., & Ben-Zvi, D. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 3-15). Dordrecht: Kluwer Academic Publishing.
- Garfield, J. B., Hogg, B., Schau, C., & Whittinghill, D. (2002). First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education* 10 (2). [Online: www.amstat.org/publications/jse/v10n2/garfield.html].
- Harris, M. B., & Schau, C. (1999). Successful strategies for teaching statistics. In S.N. Davis, M. Crawford, & J. Sebrechts (Eds.), *Coming into her own: Educational success in girls and women* (pp. 193-210). San Francisco: Jossey-Bass.
- Hau, K. T., & Marsh, H. W. (2004). The use of item parcels in structural equation modelling: Non-normal data and small sample sizes. *British Journal of Mathematical Statistical Psychology*, 57, 327-351.
- Hilton, S. C., Schau, C., & Olsen, J. A. (2004). Survey of Attitudes Toward Statistics: Factor structure invariance by gender and by administration time. *Structural Equation Modeling*, 11 (1), 92-109.
- Jolliffe, F. (1998). What is research in statistical education? In L. Pereira-Mendoza, L. Seu Kea, T. Wee Kee, & W.K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 801-806). Singapore: International Statistical Institute.

- Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kline, R. B. (2005). *Principles and practice of structural equation modelling* (2nd ed.). New York, NY: Guilford Press.
- Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6, 59-98.
- Liu, H. J. (1998). *A cross-cultural study of sex-differences in statistical reasoning for college students in Taiwan and the United States*. Doctoral dissertation, University of Minnesota, Minneapolis.
- Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33, 181-220.
- McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning, a project of the National Council of Teachers of Mathematics* (pp. 575-596). New York: Macmillan.
- Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 17-46). Dordrecht: Kluwer Academic Publishing.
- Rumsey, Deborah J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, 10(3).
[Online: <http://www.amstat.org/publications/jse/v10n3/rumsey2.html>].
- Schau, C. (2003). *Students' attitudes: the "other" important outcome in statistics education*. Paper presented in the Joint Statistical Meetings, San Francisco, CA.
- Schau, C., Dauphinee, T. L., Del Vecchio, A., & Stevens, J. (1999). *Survey of attitudes toward statistics (SATS)*.
[Online: <http://www.unm.edu/~cschau/downloadsats.pdf>].
- Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The Development and Validation of the Survey of Attitudes Toward Statistics. *Educational and psychological measurement*, 55(5), 868-875.
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.
- Sorge, C., & Schau, C. (2002). *Impact of engineering students' attitudes on achievement in statistics*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Sundre, D.L. (2003). *Assessment of Quantitative Reasoning to Enhance Educational Quality*, AERA annual meeting presentation, Chicago.
[Online: http://www.gen.umn.edu/artist/articles/AERA_2003_QRQ.pdf].
- Tempelaar, D. T. (2004). Statistical reasoning assessment: an Analysis of the SRA instrument. *Proceedings of the ARTIST Roundtable Conference*, Lawrence University. [Online: <http://www.rossmanchance.com/artist/Proctoc.html>].
- Tempelaar, D. T., Gijssels, W. H., & Schim van der Loeff, S. (2006). Puzzles in Statistical Reasoning. *Journal of Statistics Education*, 14(1).
[Online: <http://www.amstat.org/publications/jse/v14n1/tempelaar.html>]

- Tempelaar, D. T., Gijselaers, W. H., Schim van der Loeff, S., & Nijhuis, J. (2007). A structural equation model analyzing the relationship of student achievement motivations and personality factors in a range of academic subject-matter areas. *Contemporary Educational Psychology*, 32(1), 105-131.
- Tempelaar, D. T., Schim van der Loeff, S., Gijselaers, W. H., & De Crombrughe, D. (2007). *Preferred learning approaches and statistical reasoning*. Manuscript in preparation.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy -- value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68-81.
- Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Eds.), *Development of Achievement Motivation* (pp. 92-122). San Diego: Academic Press.

Chapter 4

COMMONALITIES IN ATTITUDES AND BELIEFS TOWARD DIFFERENT ACADEMIC SUBJECTS*

1. INTRODUCTION

The growth of student-centered learning as a paradigm in education at the expense of teacher-centered learning, has caused an important shift in research in learning processes: from the investigation of the impact of different instructional regimes, to the research of the influence of student characteristics and learning context on the learning process. An important element in this type of research refers to the interaction of student characteristics and context: different students perceive one and the same learning context in different ways, and develop, on the basis of these perceptions, different approaches to learning. This interaction is the focus of the so-called process stage in Presage-Process-Product or 3P models of learning.

An intriguing issue in developing 3P models of learning is the demarcation of Presage versus Process: what students' characteristics are so stable that they are dominantly outside the control of teachers, curriculum developers, student advisors or other staff (and thus belong to the Presage), and what students' characteristics are, to an important extent, influenced by contextual factors (thus belonging to the Process). In this

* This chapter is based on: Tempelaar, D. T., & Nijhuis, J. F. H. (2007). Commonalities in attitudes and beliefs toward different academic subjects. In: M. K. McCuddy, H. van den Bosch, J. W. B. Martz, A. V. Matveev, & K. O. Morse (Eds.), *Educational Innovation in Economics and Business X: The challenges of educating people to lead in a challenging world* (pp. 225-250). Berlin: Springer.

contribution, we focus on a specific but important aspect of student characteristics: attitudes and beliefs of students toward different subjects. Most research studies investigating the role of academic affective factors on learning have concluded that students' affect toward different subjects demonstrates strong variability, indicating that these affects should be regarded as Process variables (see e.g., Marsh & Yeung, 1996). However, in most of these studies self-concept is an important ingredient of the affective variables that have been investigated. Since self-concept is known to be dependent upon students' success in earlier learning experiences, the variability of affect scores is more or less implied by variation in earlier learning experiences. Focusing on a more restricted set of affective factors, attitudes and beliefs toward different subjects may increase the perspective of finding commonalities between subjects.

The study reported in this chapter reflects students who attend the problem-based program of International Business Study at the University of Maastricht. Over a period of two years, several questionnaires were administered, each measuring attitudes and beliefs to different academic subjects: statistics, business strategy, finance and accounting, marketing management, and organization. These parallel questionnaires were all based on the SATS instrument, the 'Survey of Attitudes Toward Statistics,' that was adapted to the several subjects. From a methodological point of view, data from these questionnaires were used to create a latent variables attitudes model that not only allows investigation of the existence of commonalities between subject affects, but also determination of the size of these common elements relative to the size of subject-specific affects.

2. METHODOLOGICAL BACKGROUND OF THIS STUDY

2.1 Modeling student learning

The 'Presage-Process-Product modeling approach of classroom learning' was developed by Biggs (1993, 1999) and elaborated by Prosser and Trigwell (1999). According to this 3P modeling approach, learning is seen as a progression from Presage (teaching context) through Process (teaching acts) to Products (class achievement). The approach distinguishes several building blocks in explaining learning outcomes (see Figure 4.1, adapted from Prosser & Trigwell, 1999).

The Presage building block containing student-based factors develops from an individual differences perspective in psychology on student learning. Relevant factors include abilities, prior knowledge, motivation, personality facts, learning styles, stabilized learning approaches, and so on. The second building block builds on traditional staff-developmental models, and focuses on teacher behavior. The third building block, strategies for handling the task, derives from information processing psychology, focuses on the process, the efficiency with which basic cognitive strategies are developed. The complete model integrates teaching-based, student-based, and process-based approaches to learning, and in addition to that, allows for feedback from Process to Presage, and from Product to Process and Presage. Since feedback and feedforward processes are rather different in nature, an important issue, both from a research and an applied educational point of view, is the demarcation of the several building blocks: “What, for example, belongs to students’ characteristics, and is to be regarded as rather stable and independent of the context, and what belongs to students’ perceptions of this context?”

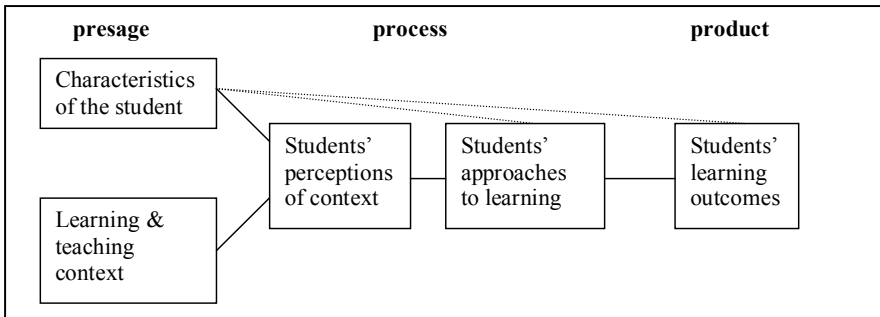


Figure 4-1. 3P or Presage-Process-Product Model of Classroom Learning

A related perspective on this demarcation issue can be cast into the ‘aptitudes for learning’ approach developed by Snow (Snow, Corno, & Jackson, 1996; Snow & Jackson, 1993). According to this approach, human learning is determined by a wide range of psychological constructs. A common classification of these constructs is the dichotomy of personality versus intelligence; however, Snow advises avoiding these ‘cloudy concepts as too complex and vague’. Instead, he proposes distinguishing among the three modes of mental functioning: cognition, conation or volition, and affection. Each of these modes is primarily defined by an

enumeration of the several processes it includes; see Figure 4.2, adapted from Snow, Corno, and Jackson (1996). As a first step of this enumeration, the three modes can be subdivided into two underlying sub-domains for each mode: temperament and emotion for affection; motivation and volition for conation; and procedural and declarative knowledge for cognition. As a second step, Figure 4.2 lists the constructs that constitute the sub-domains.

According to this taxonomy, attitudes can be of different kinds: some are (more) affective in nature, others (more) conative. Similar remarks refer to beliefs and values. This leaves us with the empirical question: “Is it possible to unravel constructs like attitudes into these different constituent parts?”

‘Personality’		‘Intelligence’			
Affection		Conation		Cognition	
Temperament	Emotion	Motivation	Volition	Procedural knowledge	Declarative knowledge
General and special personality factors		Achievement orientations	Action Controls	Skills	Domain knowledge
Values		Self-directed orientations	Other-directed orientations	Strategies Tactics	
	Attitudes	Values, Attitudes, Interest, Beliefs	Personal Styles		Beliefs

Figure 4-2. Taxonomy of Individual Difference Constructs in Aptitudes for Learning Approach, Adapted from Snow, Corno, and Jackson (1996)

2.2 Subjects of the study

Active learning, or the related concept of student-centered learning, can be implemented through different educational methods. Problem-based learning is one manifestation of student-centered learning and the one adopted at the University of Maastricht (UM) with its several faculties and schools since the school’s founding in the late 1970s. In 1984 the Faculty of Economics was established as part of the University of Maastricht. The faculty offers English-language International Business and Economics programs, as well as a Dutch one. All programs are characterized by a student-centered educational approach known as problem-based learning.

Students develop their knowledge and skills while working on problems in small groups. This is the dominant instructional mode, and only in a limited number of service courses are these tutorial sessions supplemented with more traditional lectures. The first-year service courses in Quantitative Methods (QM) are an example of such a hybrid educational approach. The three QM courses cover subjects from mathematics, statistics, and computer skills. The material is often regarded as being difficult and unattractive by most of our students.

QM is a series of required service courses for our students; its' nature of being obligatory is certainly responsible for most of the audience. The 'more attractive' courses from the perspective of our students are without doubt the business and economics courses; they really deal with the topics students expect in a business school. Table 4.1 contains the undergraduate program of the study of International Business. As is apparent from the structure, the program is organized according to a semester system with four 8-week courses per semester, always two in parallel. Courses printed in bold are the courses for which students' attitudes toward the subject were investigated.

Table 4-1. Undergraduate program of International Business Study

Courses in Undergraduate Study of International Business		
Semester 1, Block 1	Organization & Marketing	Quantitative Methods 1
Semester 1, Block 2	International Business	Quantitative Methods 2
Semester 2, Block 3	Economic & Social Sciences	Quantitative Methods 3
Semester 2, Block 4	Finance & Accounting I	Financial Inf. Systems
Semester 3, Block 1	Business Strategy	International Economics
Semester 3, Block 2	Finance & Accounting II	Quantitative Methods 4
Semester 4, Block 3	Marketing Management	Business Informatics
Semester 4, Block 4	International Business Law	Organization & HRM

Of the approximately 1,000 students participating in first-year courses, a great majority are first-year students (900). The remaining students are 'repeat' students who did not manage to pass that specific course the year before. The 900-student inflow can be decomposed according to nationality: about 600 Dutch students, 240 German students, and 60 students of other nationalities. That decomposition is important since huge differences exist between secondary school systems in Europe. All Dutch first-year students entering our school having participated in a final, national exam in at least seven subjects, including either basic mathematics

(calculus oriented), or advanced mathematics (algebra and geometry oriented), or both. In contrast, German pupils have four subjects in their final exam: two at an advanced level, two at a basic level. Having chosen mathematics in their final exam ('Abitur'), either at advanced level ('Leistungskurs') or at basic level ('Grundkurs'), their mathematical schooling is somewhat comparable to that of Dutch students in the two different groups. However, a sizeable proportion of German freshmen did not select mathematics at any level for their final exam, and their level of mathematical schooling is of a really different order compared to that of Dutch students. Besides that, the share of statistics and probability theory in mathematical courses will differ from state to state in Germany, while in the Netherlands those two topics are incorporated in basic mathematics (preparing for social sciences), but not in advanced mathematics (preparing for sciences).

With regard to prior knowledge in economics and business, similar differences exist. Students with a Dutch educational background have mostly taken introductory courses in business, or economics, or both. Inflow with a foreign diploma, in general, does not have the opportunity to achieve any prior education in these two subjects.

2.3 Students' attitudes and beliefs toward academic subjects

In the context of mathematics education, the study of affective factors in learning processes has a long tradition and has given rise to terms like 'mathematics anxiety' that seem to be reserved for the mathematics domain only. In conceptualizing the affective domain of mathematics education, McLeod (1992) distinguishes between emotions, attitudes, and beliefs. Emotions are fleeting positive and negative responses triggered by one's immediate experiences while studying mathematics. Attitudes are relatively stable, intense feelings that develop as repeated positive or negative emotional responses are automated over time. Beliefs are individually held ideas about mathematics, about oneself as a learner of mathematics, and about the social context of learning mathematics that together provide a context for mathematical experiences. In many studies of learning processes, the focus is on beliefs and attitudes, rather than emotions, which are transient and hard to measure directly, but serve as a source for the development of attitudes and are thus measured indirectly (see, for example, Gal & Garfield, 1997). A large body of literature on the role of

attitudes and beliefs toward statistics has developed, borrowing ideas from the research in mathematics education, in which one question keeps reappearing: "Are attitudes and beliefs toward learning statistics distinct from the more general ones, such as toward learning mathematics, or toward exams in general?" Gal and Ginsburg (1994) and Gal and Garfield (1997) are examples of this line of research. The issues investigated in this research are a natural continuation of this question: "Are attitudes and beliefs toward statistics distinct from other attitudes and beliefs, such as those toward business subjects for students studying business?"

The area of research on developing instruments to assess attitudes and beliefs toward statistics is well developed. In the 1980s, several instruments were developed, all using statements for which respondents mark their agreement or disagreement on 5-point or 7-point, Likert-type scale. These include the Statistics Attitude Survey (see Roberts & Bilderback, 1980; Roberts & Saxe, 1982), the Statistical Anxiety Rating Scale (see Cruise, Cash, & Bolton, 1985), the Statistical Anxiety Inventory (see Zeidner, 1991), and the Attitudes Toward Statistics (see Wise, 1985). As each of these instruments had some drawbacks, Schau, Stevens, Dauphinee, and DelVecchio (1995) developed the Survey of Attitudes Toward Statistics (SATS) in the 1990s. In our research, we opted for the SATS instrument on the grounds of the theoretical reasons that led to its development and the fact its statistical properties are well documented.

In their review article, Gal and Garfield (1997) distinguish several reasons to take affective factors into account. Attitudes and beliefs about statistics influence the learning and teaching of statistics, and the willingness of students to enroll in elective statistics courses. There is, however, more than this process consideration. Amongst the goals of education, the development of problem-solving capabilities, literacy and related communication skills, and data-analyzing skills becomes increasingly important. To reach those goals, it is important to remove negative attitudes and beliefs – and in doing that, we should assess and monitor students' attitudes.

Gal and Garfield (1997) distinguish among three different sources for attitudes and beliefs toward statistics: previous experience with statistics in school-related contexts, 'notions' on what statistics means based on out-of-school-lives, and attitudes toward mathematics that are merely transferred to statistics. A proper definition of the several aspects of students' ideas, feelings and reactions about academic subjects, and the learning of them, is, however, still a challenging issue. As Gal and Garfield (1997, p. 40)

remark, researchers in statistical education have used the terms ‘attitudes’ and ‘beliefs,’ and also the linkage of the two, without paying much attention to an explicit definition of the terms and the distinction between them. In applying McLeod’s terminology to education in statistics and several business subjects, we will follow Gal and Garfield (1997) in focusing on beliefs and attitudes, rather than emotions, which are transient and hard to measure directly, but serve as a source for the development of attitudes and are thus measured indirectly. Beliefs are relatively stable and resistant to change; it takes time to develop them, and cultural factors play an important role in their development. They possess a larger cognitive component and less emotional intensity than attitudes. Attitudes toward an academic subject represent an accumulation of emotions and feelings experienced over time in the context of learning that and related subjects. They are rather stable with moderate intensity, having a smaller cognitive component than beliefs. Attributes are expressed along a positive-negative continuum, and may represent feelings, for example, toward a topic or activity.

An examination of research on learning business subjects suggests that, although the role of student attitudes in learning has been investigated, these attitudes do not refer to specific business subjects, but instead to more general aspects. Examples are research of students’ attitudes toward academic group work, toward different instructional formats, toward school or university in general, or to instructional tools as computers (see for example, Gardner & Korth, 1998; Ruggiero, 1998).

2.4 SATS: Survey of Attitudes Toward Statistics

To assess student attitudes and beliefs regarding statistics, Schau et al. (1995) developed the Survey of Attitudes Toward Statistics (SATS). The SATS contains 28 items to identify attitudes about statistics. The SATS scales, each accompanied by two examples of items, one positively and one negatively worded, are (see Gal & Garfield, 1997, p. 44):

- Affect: measuring positive and negative feeling concerning statistics: I like statistics; I am scared by statistics.
- Cognitive Competence: measuring attitudes about intellectual knowledge and skills when applied to statistics: I can learn statistics; I have no idea of what’s going on in statistics.
- Value: measuring attitudes about the usefulness, relevance, and worth of statistics in personal and professional life: I use statistics in

my everyday life; I will have no application for statistics in my profession.

- Difficulty: measuring attitudes about the difficulty of statistics as a subject: Statistics formulas are easy to understand; Statistics is highly technical.

The SATS has two forms, with minor differences in wording: a 'pre' form for students who have not yet taken a statistics course, and a 'post' form for administration during or after a course. Since most students were already familiar with the subject statistics, it was decided to administer the post version.

In total, attitudes and beliefs inventories based on the SATS instrument were administered for five different subjects (see also Table 4.1): Statistics, International Business Strategy, International Finance and Accounting II, International Marketing Management, and International Organization and Human Resource Management (HRM) (since all courses in the International Business studies have 'International' as first part of the their name, it will be skipped in all further references). The Statistics questionnaire was administered in the very first week of the first semester; the questionnaires for the several business subjects were administered in the third and fourth semester (see once again Table 4.1 for the exact timing). The Statistics SATS, being part of a larger collection of inventories, was administered using a 5-point, Likert- type scale, contrary to the original SATS. The practical reason to do so was based on the wish to have one item format for all entry questionnaires. Anchors were, however, the same: 'strongly disagree' (left-hand anchor), 'neither disagree nor agree' (center anchor), and 'strongly agree' (right-hand anchor).

All 984 students in the first-year programs of the Economics studies and International Business studies participated in the administration of the SATS Statistics in the first semester. Since the investigation of the robustness of SATS outcomes over several subgroups (see next section) was one of the aims of the analysis, it was decided to keep all students in the sample. Based on the notion of robustness over subgroups, we can anticipate the empirical results of the next section: outcomes for both studies will be similar (for all variables, the hypothesis of invariance over studies will not be statistically rejected), justifying the assumption that characteristics of students from both studies can be regarded as equal.

3. ATTITUDES AND BELIEFS TOWARD STATISTICS

As a prelude to our empirical results, we will summarize findings in other reported applications. Schau et al. (1995) and Gal and Garfield (1997) report that scores on the four scales vary in their interrelationship. Scores on Affect and Cognitive Competence are strongly related; scores on Value and Difficulty are moderately related to Affect and Cognitive Competence but unrelated to each other. Internal consistencies of all scales were found to be adequate, ranging from .6 to above .8 (see for example, Hair, Anderson, Tatham, and Black, 1998, for a discussion of cut-off points). In a confirmatory analysis comparing factor models of dimensions 1 to 4, Dauphinee, Schau, and Stevens (1997) concluded that the four-factor model had the best statistical properties, both in an absolute and relative (that is, corrected for degrees of freedom) sense; however, improvement over a three-factor model was limited (we will return to this issue in the next section). In the four-factor structural equation model developed in Dauphinee, Schau, and Stevens (1997), the factors are not based upon separate items but on two to three item parcels or testlets. 'Item parcelling' is the technique, applied in many large scale SEM studies, to replace dozens of items all loading on one factor, by a restricted number (two to four) parcels or mini-scales loading on that factor. Each parcel is found by aggregating related items. Item parcelling has several advantages, such as reducing the dimension of the factor model, and improving the reliability of the measured variables. We will refer to Dauphinee, Schau, and Stevens (1997) once again, this time not for its discussion of SATS, but as one of the prime references of item parcelling in applied educational psychology. We apply item parcelling in our study, and base the choice of parcels on the ones developed in Dauphinee, Schau, and Stevens (1997).

Descriptive statistics for several subgroups of students in the studies Economics and International Business are contained in Table 4.2. All mean scores for Affect (AFFECT), Cognitive Concept (COGNC), and Value (VALUE) are far above (and statistically significantly different from) the neutral level of three: students of different background have positive attitudes and beliefs in these aspects (of 32 hypothesis tests, 30 result in p -values smaller than .0005, the other two in .006 and .039). In contrast, all mean scores for Difficulty (DIFFIC) are below the neutral level, expressing that students perceive the subject as difficult (the naming of the Difficulty-scale is somewhat counter intuitive: all scales are defined such that higher values correspond to more positive attitudes and feelings; a name like 'lack

of perceived difficulty' would better catch this meaning). The table demonstrates that students can evaluate themselves as rather cognitively competent in learning statistics (mean score of 3.52 on the 5-point scale), and at the same time regard statistics as a somewhat difficult topic.

Table 4-2. Average scores for scales of Affect, Cognitive Competence, Value, and Difficulty in several categories of students

	Dutch UM students			German UM students			All	All
	Female (n=204)	Male (n=353)	Total (n=557)	Female (n=92)	Male (n=125)	Total (n=217)	Female (n=386)	Male (n=560)
AFFECT	3.26	3.49	3.40	3.14	3.24	3.19	3.19	3.39
COGNC	3.36	3.65	3.54	3.47	3.61	3.55	3.38	3.62
VALUE	3.66	3.70	3.68	3.81	3.78	3.79	3.69	3.71
DIFFIC	2.68	2.80	2.76	2.56	2.62	2.59	2.62	2.74

Table 4.2 suggests that both gender and nationality effects may be present. Performing independent samples *t*-tests confirms this impression: male students have significantly higher scores on Affect, Cognitive Competence, and Difficulty (all *p*-values less than .001, given the very small values of all standard errors: in the range of 0.01 to 0.05); for Value, no difference exists. In comparing Dutch and German students, once again three scales demonstrate significant differences. German students score significantly higher in Value, but significantly lower in Affect and Difficulty than Dutch students, while the score on Cognitive Concept is invariant across nationalities. Figure 4.3 illustrates these patterns of gender and nationality effects in a graphical way, demonstrating several 95% confidence intervals for scale means. In this graph, all students with a non-Dutch secondary diploma (most of them of German nationality) are integrated in one category.

With regard to the correlation structure of the four attitudes scales, our findings support the results reported in Gal and Garfield (1997): Affect and Cognitive Competence are strongly related; Value and Difficulty are moderately related to Affect and Cognitive Competence but are not interrelated. See Table 4.3 for the correlations between the several scales for all students.

The internal consistency is evaluated by calculating the Cronbach alpha coefficients for internal consistency; their values are also given in Table 4.3.

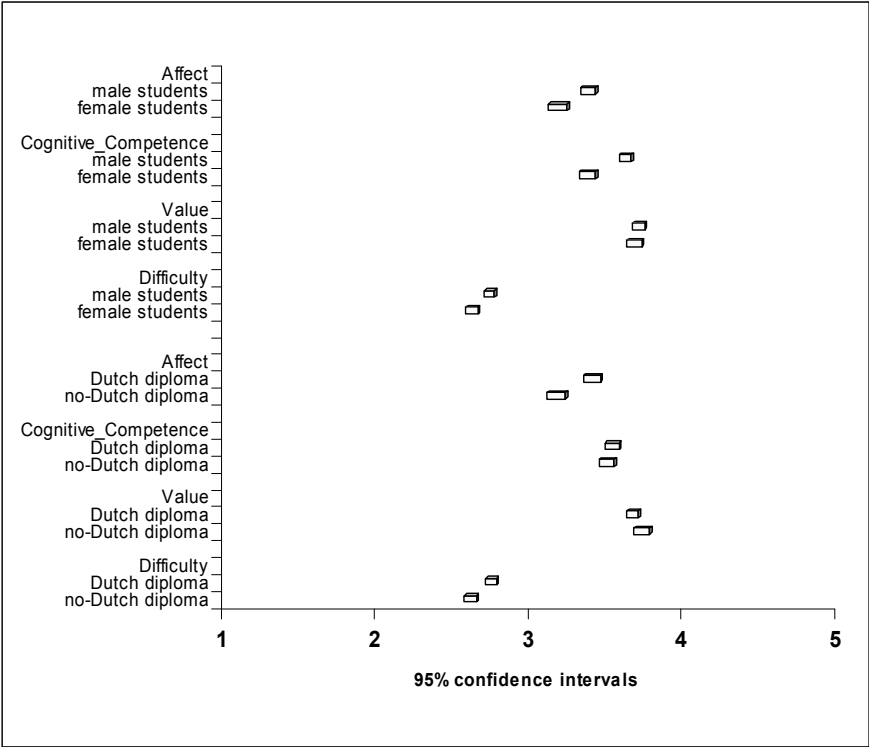


Figure 4-3. 95% Confidence intervals of mean attitude scores for male versus female students and students with a Dutch diploma versus students with a non-Dutch diploma

Table 4-3. Correlations between the four attitude scales, and Cronbach alpha coefficients for internal consistency with reference values from Schau et al. (1997)

	Correlations				Cronbach Alpha Reliabilities	
	AFFECT	COGNC	VALUE	DIFFIC	This Study	Schau <i>et al.</i>
AFFECT	1.00				.80	.81 - .85
COGNC	.70	1.00			.72	.77 - .83
VALUE	.35	.38	1.00		.76	.80 - .85
DIFFIC	.50	.49	.10	1.00	.62	.64 - .75

In the last column, reference values are given based upon Schau et al. (1995) and Dauphinee, Schau, and Stevens (1997). The values given in that column specify a range derived from several studies and samples. The

alpha coefficients of all scales are, although somewhat lower than the reference values, satisfactory. Not removing extreme cases from the data set, as in some other studies, may be a possible explanation for the differences compared to the reference values.

Several studies of gender differences in attitudes toward statistics are reviewed in Dauphinee, Schau, and Stevens (1997). Although the general conclusion of all these studies tends to be the same, there are some restrictions in comparing their outcomes since different researchers used different instruments, each having different scales. Roberts and Sache (1982), using the one-dimensional Statistics Attitude Survey developed by Roberts and Bilderback (1980), concluded that male students exhibit, on average, more positive attitudes than female students, both at the beginning and at the end of an introductory statistics course. This is, in short, also the general conclusion of other research, but then with much more nuance brought forward by the use of multi-dimensional attitude scales. In an application of Wise's (1985) Attitudes Toward Statistics, Waters, Martelli, Zakrajsek, and Popovich (1988) found that male students have more positive course attitudes than female students, whereas no gender differences exist with respect to the other attitude scales. These outcomes resemble our conclusions as reported in Table 4.3. Dauphinee, Schau, and Stevens (1997) study a different aspect of gender effects: not the invariance of means, but the invariance of factorial structures. We will extend our study in that direction in the next paragraph, in which a measurement model of attitudes is discussed.

As a preliminary step to confirmatory factor analysis on the SATS-data, we proceed with the analysis by 'item parcelling': forming parcels or mini-scales by aggregating item scores. Nine item-parcels are formed in exactly the same way as described in Dauphinee, Schau, and Stevens (1997): two parcels for Affect (AFFPARC1 and AFFPARC2), two parcels for Cognitive Concept (COCPARC1 and COCPARC2), three parcels for Value (VALPARC1, VALPARC2, and VALPARC3), and two parcels for Difficulty (DIFPARC1 and DIFPARC2). Based on these parcels, a measurement model for four latent variables AFFECT, COGNC, VALUE and DIFFIC was estimated in different subsamples: females and males, and Dutch and German students.

The structure of these models appeared to be extremely stable: no cross-loading emerged (in more than one subsample), and all latent factors, except for VALUE and DIFFIC, have significant positive correlations. The correlation between VALUE and DIFFIC is also positive, but not

significantly in all samples, and thus was restricted to be zero. The final model, estimated on the total sample, is depicted in Figure 4.4 and Table 4.4.

The structure of this final model is identical to that of the final model described in Dauphinee et al. (1997), confirming the robust character of the SATS inventory. To allow an impression of the similarities of both models, the correlation structure of latent errors is reproduced from Dauphinee et al. (1997) in Table 4.4.

Figure 4-4. Final measurement model of attitudes and beliefs:
Factor model with unstandardized factor loadings

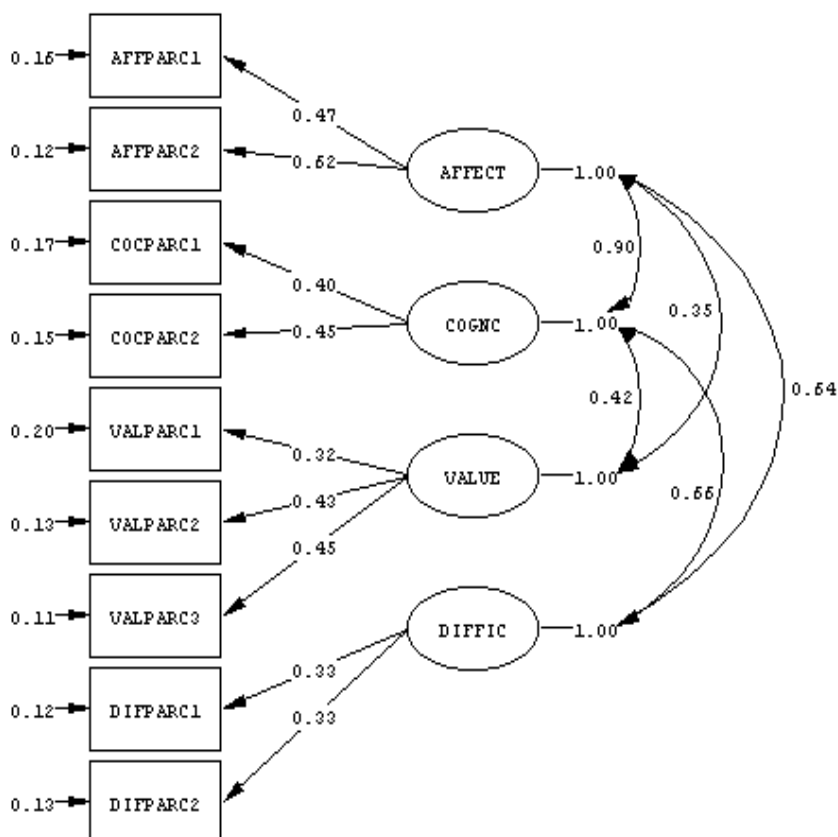


Table 4-4. Final measurement model of attitudes and beliefs: Completely standardized factor loadings and correlation of latent errors (including reference values)

Completely standardized factor loadings				
	AFFECT	COGNC	VALUE	DIFFIC
AFFPARC1	0.76			
AFFPARC2	0.87			
COCPARC1		0.70		
COCPARC2		0.77		
VALPARC1			0.58	
VALPARC2			0.76	
VALPARC3			0.80	
DIFPARC1				0.69
Correlation matrix of latent errors (reference values, in brackets, from Dauphinee et al. (1997))				
	AFFECT	COGNC	VALUE	DIFFIC
COGNC	0.90 (0.94)			
VALUE	0.35 (0.32)	0.42 (0.39)		
DIFFIC	0.64 (0.73)	0.66 (0.64)	- (-)	

We continue the analysis by investigating gender and nationality effects. We perform sequentially more restrictive tests of model invariance as in Dauphinee et al. (1997). In that article a factor structure and factor loadings were found that were invariant across gender, but the hypothesis of invariance with regard to factor variances and covariances had to be rejected. The latter was caused by females having a much higher variance in Value scores, and a much smaller correlation between Value and Affect scores, than males. In our sample, we were not able to reproduce these empirical findings. In testing for a nationality effect, the last and most restrictive test was significant, indicating that the hypothesis of invariant factor structure and error correlations cannot be rejected, but that the variability in scores on the several parcels is different for Dutch and German students. Differences appeared to be in the Value and Difficulty parcels, the variability being much larger in German students than in Dutch students.

A very last step in testing for gender and nationality effects is to perform a test on latent means. Dauphinee et al. (1997) do not report on this type of test, so no reference values are available. In testing for gender effect, we find large differences for three of the four latent variables: for the latent variables AFFECT, COGNC, and DIFFIC, the means of male students are significantly higher than that of female students (p -values all smaller than

.001), while for the remaining latent variable, VALUE, no significant difference in means exist. For the nationality case, the following pattern arises: in comparison to German students, Dutch students score significantly higher in AFFECT and DIFFIC, significantly lower in VALUE, and similarly in COGNC. These outcomes are similar to the ones achieved in performing independent samples t-test on scale means.

These conclusions seem to add to the validation of the SATS-instrument and the factor model based on that instrument. In contrast to findings in Dauphinee et al. (1997), the model created on the whole sample appears to be valid for all relevant subsamples, even though these subsamples show large differences in characteristics, and even though our sample is extremely large (in comparison to other studies). However, these differences are best interpreted as differences in latent means, while the factorial structure is invariant. This leaves the question of: "Why are there large differences in the means?" The difference in means across nationalities can easily be explained by the better match between secondary and higher education. The introduction to statistics in Dutch secondary schools could explain the higher scores in Affect and Difficulty of Dutch students, while the strong study orientation of German students explains their high Value score. However, we are at a loss to provide a good explanation behind the large gender effect.

4. ATTITUDES AND BELIEFS TOWARD BUSINESS SUBJECTS AND STATISTICS

One year after the administration of the SATS Statistics, attitudes and beliefs toward four different international business subjects were assessed for a subset of the students. This subset consists of students in the study of International Business (so excluding Economics students) who passed their propedeutic exam (in the Dutch tertiary education system, the first year program is finished with an exam, called the propedeutic exam, that serves as an entry exam for the second year program; students failing this propedeutic exam redo the first year program as 'repeat students'). To assess students' attitudes and beliefs toward Business Strategy, Finance and Accounting, Marketing Management, and Organization and Human Resource Management (HRM), the SATS inventory was adapted by replacing the subject name Statistics by each of the four subject names mentioned. In addition to that, two items were slightly reformulated, to

accommodate the different nature of business subjects. The statement 'Statistics formulas are easy to understand' was rephrased as 'Techniques of Business Strategy are easy to understand,' and the statement 'Statistics involves massive computations' as 'Business Strategy involves a lot of technical aspects' (similar for the other subjects). The inventories were administered using the original 7-point Likert scale; to make the outcomes comparable to the Statistics outcomes, all scores were re-expressed in terms of a 1-5 scale. This re-expression using a linear transformation was performed so as to pertain to the three anchors (so $1 \Rightarrow 1$, $4 \Rightarrow 3$, and $7 \Rightarrow 5$, preserving left-hand, centre, and right-hand anchors). To eliminate any confounding factors in the comparison between subjects, the data on which the Statistics attitudes and beliefs are based are restricted to those students for which data is available for the other courses. Table 4.5 contains descriptive statistics for all four attitudes and beliefs scales in all five subjects, including the relevant sample size.

Table 4-5. Means, and standard errors per attitude scale per subject, and sample size n , for subjects: Statistics (1.1QM), Business Strategy (2.1BS), Finance and Accounting (2.2FA), Marketing Management (2.3MM), and Organization and Human Resource Management (2.4OHRM)

Attitude Scales	Affect		Cognitive Competence		Value		Difficulty		Size
	Mean	St.err.	Mean	St.err.	Mean	St.err.	Mean	St.err.	N
Subjects									
1.1QM	3.23	0.03	3.47	0.03	3.70	0.03	2.62	0.02	346
2.1BS	4.02	0.04	4.09	0.04	4.04	0.04	3.24	0.04	349
2.2F&A	3.00	0.06	3.29	0.05	3.54	0.04	2.23	0.04	320
2.3MM	3.95	0.05	3.96	0.04	3.97	0.05	3.17	0.04	267
2.4OHRM	3.91	0.04	3.87	0.04	3.81	0.04	3.35	0.04	281

The descriptive statistics displayed in Table 4.5 show a clear pattern. For the subjects Business Strategy, Marketing Management, and Organization and HRM, students express positive attitudes and beliefs: all scale means are far (and statistically significant given the small values of the standard errors; all t -values of testing against the neutral level are larger than 4.3, all p -values smaller than .0005) above the neutral level of 3. On top of that, differences in scale means between the three subjects are very small. In contrast, the descriptives for both Statistics and Finance and Accounting are much less positive, and even negative for the scale Difficulty. Although these two subjects, perceived as the most analytical subjects of the

program, both have less positive attitudes, scores are not identical: Finance and Accounting tends to have an even tougher reputation than Statistics.

Since LISREL estimates are based upon the correlation structure of the several items, a further impression can be achieved by studying the correlation structure of the several attitude scales per subject. These are expressed in Table 4.6.

Table 4-6. Correlations between attitude scales per subject, for subjects: Statistics (1.1QM), Business Strategy (2.1BS), Finance and Accounting (2.2F&A), Marketing Management (2.3MM), and Organization and Human Resource Management (2.4OHRM).

1.1QM	AFFECT	COGNC	VALUE	DIFFIC
AFFECT	1.00			
COGNC	0.71	1.00		
VALUE	0.35	0.40	1.00	
DIFFIC	0.53	0.54	0.13	1.00
2.2F&A	AFFECT	COGNC	VALUE	DIFFIC
AFFECT	1.00			
COGNC	0.78	1.00		
VALUE	0.47	0.49	1.00	
DIFFIC	0.50	0.52	0.10	1.00
2.1BS	AFFECT	COGNC	VALUE	DIFFIC
AFFECT	1.00			
COGNC	0.62	1.00		
VALUE	0.54	0.50	1.00	
DIFFIC	0.30	0.44	0.05	1.00
2.3MM	AFFECT	COGNC	VALUE	DIFFIC
AFFECT	1.00			
COGNC	0.63	1.00		
VALUE	0.50	0.49	1.00	
DIFFIC	0.27	0.43	0.01	1.00
2.4OHRM	AFFECT	COGNC	VALUE	DIFFIC
AFFECT	1.00			
COGNC	0.60	1.00		
VALUE	0.61	0.36	1.00	
DIFFIC	0.26	0.55	-0.03	1.00

Table 4.6 confirms the similarity of the three management-oriented subjects, in contrast to the two analytical subjects. Two striking differences are: the correlation between Value and Difficulty, the smallest correlation of the four attitudes scales for Statistics, completely disappears in the right-hand side of Table 4.6: for the management-oriented subjects, Value and Difficulty are unrelated concepts. And the very high correlation between Affect and Cognitive Competence for both analytical subjects gets much smaller for the management-oriented subjects, whereas the correlation of Affect and Value rises for the later subjects.

The next step in the statistical analysis of the subject attitudes is the estimation of a four-factor, first-order confirmatory factor model for each of the subjects separately. For all subjects, the same model structure was imposed: factors were allowed to covary, except for attitudes Affect and Value, for which the covariation was set to zero. Outcomes of the estimation of these first-order models are summarized in Tables 4.7 and 4.8, containing factor loadings and the correlation structure of the latent factors.

Table 4-7. Factor loadings of parcels for first order latent factor models per subject, for subjects: Statistics (1.1QM), Business Strategy (2.1BS), Finance and Accounting (2.2F&A), Marketing Management (2.3MM), and Organization and Human Resource Management (2.4OHRM).

1.1QM	Parcel1	Parcel2	Parcel3
AFFECT	0.81	0.87	
COGNC	0.70	0.76	
VALUE	0.56	0.80	0.86
DIFFIC	0.63	0.73	
2.2F&A	Parcel1	Parcel2	Parcel3
AFFECT	0.84	0.91	
COGNC	0.85	0.81	
VALUE	0.64	0.84	0.88
DIFFIC	0.77	0.84	
2.1BS	Parcel1	Parcel2	Parcel3
AFFECT	0.72	0.85	
COGNC	0.76	0.83	
VALUE	0.65	0.79	0.82
DIFFIC	0.76	0.79	
2.3MM	Parcel1	Parcel2	Parcel3
AFFECT	0.82	0.80	
COGNC	0.84	0.76	
VALUE	0.75	0.84	0.86
DIFFIC	0.80	0.80	
2.4OHRM	Parcel1	Parcel2	Parcel3
AFFECT	0.76	0.84	
COGNC	0.78	0.80	
VALUE	0.79	0.81	0.90
DIFFIC	0.87	0.74	

Table 4-8. Correlations between latent factors per subject, for subjects: Statistics (1.1QM), Business Strategy (2.1BS), Finance and Accounting (2.2F&A), Marketing Management (2.3MM), and Organization and Human Resource Management (2.4OHRM).

1.1QM	AFFECT	COGNC	VALUE	DIFFIC
AFFECT	1.00			
COGNC	0.91	1.00		
VALUE	0.36	0.45	1.00	
DIFFIC	0.65	0.67	-	1.00

2.2F&A	AFFECT	COGNC	VALUE	DIFFIC
AFFECT	1.00			
COGNC	0.91	1.00		
VALUE	0.48	0.53	1.00	
DIFFIC	0.59	0.62	-	1.00

2.1BS	AFFECT	COGNC	VALUE	DIFFIC
AFFECT	1.00			
COGNC	0.80	1.00		
VALUE	0.67	0.63	1.00	
DIFFIC	0.37	0.56	-	1.00

2.3MM	AFFECT	COGNC	VALUE	DIFFIC
AFFECT	1.00			
COGNC	0.81	1.00		
VALUE	0.59	0.56	1.00	
DIFFIC	0.34	0.52	-	1.00

2.4OHRM	AFFECT	COGNC	VALUE	DIFFIC
AFFECT	1.00			
COGNC	0.83	1.00		
VALUE	0.73	0.47	1.00	
DIFFIC	0.38	0.72	-	1.00

This structural equation modeling step confirms the conclusions of the descriptive analysis: judged on the basis of students' attitudes and beliefs, the program contains two kinds of subjects - analytical subjects and management-oriented subjects. Analytical subjects, such as Quantitative Methods and Finance and Accounting, give rise to attitudes where:

- The covariation between the latent factors Affect and Cognitive Competence is so strong that these attitudes nearly collapse (correlations above 0.90). This is in agreement with the finding of

Dauphinee, Schau, and Stevens (1997), where the four-factor models induced minor improvements over the three-factor model.

- The correlation between Affect and Value is relatively low: students can dislike the subject, but still regard it as highly important.

The management oriented subjects give rise to attitudes where:

- The latent factors Affect and Cognitive Competence demonstrate less covariation; the correlation is in the order of 0.8, a figure low enough to make the two factors really different.
- The correlation between Affect and Value is relatively high: the liking of the subject is a strong predictor of the attached importance of the subject, and vice versa.

5. A SECOND-ORDER FACTOR MODEL

As a last step in this modeling tour, a second-order factor model was estimated with a common second order factor for each of the attitudes, explaining the corresponding subject-specific first order attitude factors. The correlation structure of the second order factor was chosen to be identical to the correlation structures of the first order models: all attitudes were allowed to covary, except for Value and Difficulty. No cross-loadings were allowed. Figure 4.5 depicts the estimated model graphically, with the four second order latent factors on the left-hand side, explaining the 20 subject-specific first order factors in the middle of the graph, and with the 45 subject-specific item-parcels used as indicator variables on the right-hand side.

Tables 4.9-4.11 contain numerical output on the estimated second order factor model. In Table 4.9, the completely standardized factor loadings or regression coefficients that explain first order factors by the corresponding second order factors are indicated. The table demonstrates an already rather familiar pattern: factor loadings of the management-oriented subjects are different, and in this case much higher, than those of the analytical subjects.

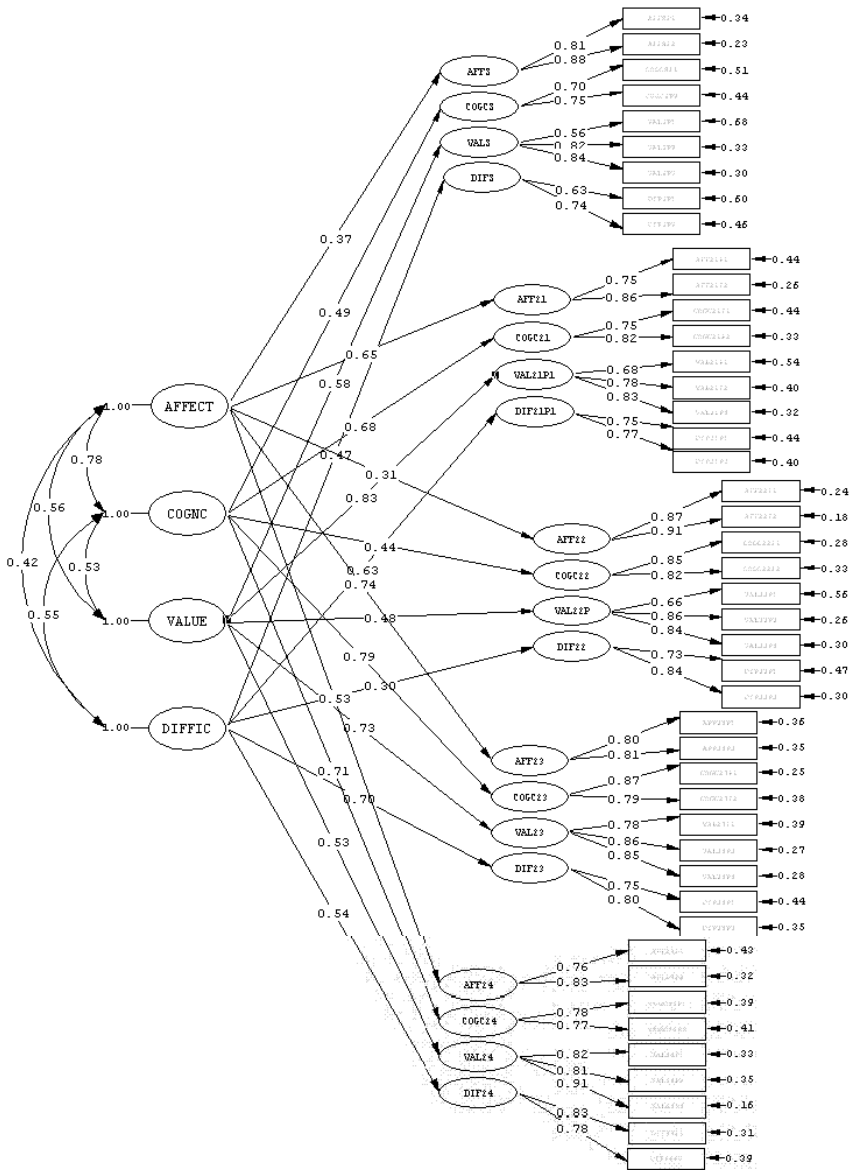


Figure 4-5. Final measurement model of attitudes and beliefs:
Factor model with unstandardized factor loadings

Table 4-9. Completely standardized factor loadings of first order factors on second order factors, for subjects: Statistics (1.1QM), Business Strategy (2.1BS), Finance and Accounting (2.2FA), Marketing Management (2.3MM), and Organization and Human Resource Management (2.4OHRM).

Subjects	AFFECT	COGNC	VALUE	DIFFIC
1.1QM	0.35	0.47	0.54	0.48
2.1BS	0.61	0.65	0.82	0.61
2.2F&A	0.31	0.45	0.45	0.37
2.3MM	0.76	0.80	0.73	0.67
2.4OHRM	0.56	0.71	0.56	0.53

Table 4.10 depicts the correlation structure of the second order factors. The values are very similar to those in the right-hand side of Table 4.8: the first order correlation structure of the factors of the management-oriented subjects.

Table 4-10. Correlations between second order latent factors

	AFFECT	COGNC	VALUE	DIFFIC
AFFECT	1.00			
COGNC	0.78	1.00		
VALUE	0.56	0.53	1.00	
DIFFIC	0.42	0.55	-	1.00

The last table, Table 4.11, provides the explained variation in the first order factors by the second order factors, expressed as the percentage value of R^2 .

Table 4-11. Explained variation of first order factors by second order factors, for subjects: Statistics (1.1QM), Business Strategy (2.1BS), Finance and Accounting (2.2FA), Marketing Management (2.3MM), and Organization and Human Resource Management (2.4OHRM).

Subjects	AFFECT	COGNC	VALUE	DIFFIC	Average
1.1QM	14	24	34	22	24
2.1BS	41	46	67	39	48
2.2F&A	9	19	23	9	15
2.3MM	56	64	53	49	56
2.4OHRM	28	50	31	30	35
Average	30	41	42	30	35

Once again, the different kinds of subjects quite clearly manifest themselves: the figures for the management-oriented subjects are quite

different from those of the analytical subjects. Another pattern appears: the explained variation in the factors Cognitive Competence and Value is much higher than that in the other two factors, Affect and Difficulty.

Summarizing all outcomes, the following picture emerges:

- There exist strong commonalities in students' attitudes and beliefs toward different subjects. Averaged over all attitudes and all subjects, somewhat more than one-third of the variation of subject-specific attitudes can be explained by deeper, subject-independent attitudes of students.
- The degree of commonality is different for diverse attitudes. The attitudes Cognitive Competence and Value have the strongest common components, irrespective of the subject. And for all subjects, the attitudes Affect and Difficulty have the strongest subject-specific components.
- Strong differences exist in the importance of the common attitudes vis-à-vis the subject-specific attitudes, between the two kinds of subjects. In the management-oriented subjects, roughly half of all variation is explained by the common component. In contrast, only about 20% of the variation in attitudes is explained by a shared attitude component, while the major part is subject-specific in the case of the analytical subjects.

6. CONCLUSIONS

From a substantive point of view, the Marsh and Yeung (1996) study aims to test two different "straw person hypotheses": that different academic affects associated with the same subject can be incorporated into a single subject-specific factor, implying that distinctions among different academic affects are not very important; and that affect in different academic subjects is quite distinct from affect in other subjects, implying that one cannot hope to describe affect with general affect factors only. Marsh and Yeung (1996) succeed in testing and rejecting both hypotheses, as we do in this study. However, Marsh and Yeung (1996, p. 684) were somewhat limited in their analysis by the restricted richness of their data set: "Unfortunately, however, data available in the NELS88 are not very strong for testing this hypothesis. Stronger tests of this hypothesis would require well-validated multi-item scales instead of single-item scales and the inclusion of a more diverse set of academic affects." All these requirements - multi-item scales, well validated, for a range of affects - are provided by the SATS and the

application of it to a diversity of subjects (be it of tertiary level, instead of secondary).

The outcomes of our study confirm those of the Marsh and Yeung (1996) study:

- Affect has two important components: one general, context independent, and one subject-specific, both of them of significant size; and
- Affect is multidimensional, that is, it is constituted of a range of different constructs.

This study, through the richness of its data, provides relevant details within this broad outline (e.g., in management-oriented subjects, four different attitudes are clearly apparent; in the analytical subjects, the attitudes Affect and Cognitive competence tend to get indistinguishable).

What is the educational implication of these findings? Contrary to the general perception of lecturers, who expect their students to have similar attitudes toward and to get similar ratings in different subjects, Marsh and Yeung (1996) find that students demonstrate different affects to different subjects. Recognition of this variation should lead to more effective teaching is the rather positive conclusion of the Marsh and Yeung (1996) study. The outcomes of our study confirm their findings: the hypothesis of invariant affects toward different subjects is to be rejected. Our study, however, also reveals a common underlying factor in these subject-specific affective factors. This conclusion that affects are a mixture of a trait-like stable component and a state-like subject-specific component has promising potential educational implications for interventions directed at adjusting negative affects. Such interventions should typically consist of a combination of interventions within courses, with the lecturer stimulating and enthusing students for specific academic disciplines, and interventions over several courses, e.g., empowering students with low self-concepts (the cognitive competence component of affects). Extreme interventions - that is only within a course or solely outside specific courses - are expected to be less efficient as interventions operating at both levels.

Further research is needed into affective factors in education. An intriguing question is the relationship between personality styles and (the common component of) attitudes toward subjects. Although intriguing, outcomes of this type of research might not offer that much to the perspectives of intervention. It is difficult to imagine (negative) attitudes being more stable than those explained by personality, and for that reason it is difficult to imagine what interventions could be successful in adjusting

those negative affects. A research question that combines both relevancy and perspectives for educational interventions refers to the relation between attitudes and student learning approaches; it is this research area on which both authors are presently working.

REFERENCES

- Biggs, J. B. (1993). From theory to practice: A cognitive systems approach, *Higher Education Research and Development*, 12, 73-86.
- Biggs, J., (1999). *Teaching for quality learning at university*. Buckingham: The Society for Research into Higher Education & Open University Press.
- Cruise, R. J., Cash, R. W., & Bolton, D. L. (1985). Development and validation of an instrument to measure statistical anxiety. *American Statistical Association Proceedings of the Section on Statistics Education*, 92-97.
- Dauphinee, T. L., Schau, C., & Stevens, J. J. (1997). Survey of attitudes toward statistics: Factor structure and factorial invariance for women and men, *Structural Equation Modeling: A Multidisciplinary Journal*, 4 (2), 129-141.
- Gal, I., & Garfield, J. B. (1997) *Curricular goals and assessment challenges in statistics education*. In I. Gal, & J. B. Garfield (Eds.), *The assessment challenge in statistical education* (pp. 1-13). Voorburg: IOS Press.
- Gardner, B. S., & Korth, S. J. (1998). A framework for learning to work in teams. *Journal of Education for Business*, 74 (1), 28-33.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate Data Analysis* (5th ed.). Upper Saddle River, NJ.: Prentice-Hall.
- Marsh, H. W., & Yeung, A. S. (1996). The distinctiveness of affects in specific school subjects: An Application of confirmatory factor analysis with the national educational longitudinal study of 1988. *American Educational Research Journal*, 33 (3), 665-689.
- McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning, a project of the National Council of Teachers of Mathematics* (pp. 575-596). New York: Macmillan.
- Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching: The experience in higher education*. Buckingham: The Society for Research into Higher Education & Open University Press.
- Roberts, D. M., & Bilderback, E. W. (1980). Reliability and validity of a statistics attitude survey. *Educational and Psychological Measurement*, 40, 235-238.
- Roberts, D. M., & Saxe, J. E. (1982). Validity of a statistics attitude survey: A follow-up study. *Educational and Psychological Measurement*, 47, 907-912.
- Ruggiero, V. R. (1998). *Changing attitudes, a strategy for motivating students to learn*. Boston: Allyn and Bacon.
- Schau, C., Stevens, J., Dauphinee, T. L., & Vecchio, A. D. (1995). The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement*, 55 (5), 868-875.

- Snow, R. E., Corno, L., & Jackson, D. (1996). Individual differences in affective and conative functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 243-310). New York: Simon & Schuster Macmillan.
- Snow, R. E., & Jackson, D. N. (1993). Assessment of conative constructs for educational research and evaluation: A catalogue. Los Angeles: CSE Technical Report 354, CRESST.
- Wise, S. L. (1985). The development and validation of a scale measuring attitudes towards statistics. *Educational and Psychological Measurement*, 45, 401-405.
- Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *British Journal of Educational Psychology*, 61, 319-328.

Chapter 5

STUDENT ACHIEVEMENT MOTIVATIONS IN BUSINESS SUBJECTS*

1. INTRODUCTION

To compete in the global economy, more engineers and scientists are needed. Indeed, the European Council has set as one of its objectives: "... to bring about a 15% increase by 2010 in the number of students in these fields ..." (Outcome of Proceedings of the Council of the European Union, February 26, 2004). In the last decade, campaigns have been staged in the Netherlands, like in many other countries, to stimulate prospective students to enroll for studies in the fields of science and technology. Such campaigns are usually predicated on vague notions such as that "smart" young people should strive for careers in "exact" fields. Evidently, these efforts are based on the idea that universal prescriptions are possible for shaping the forces that influence learners' choices. Pajares and Schunk (2005) criticized such notions, and argued in their recent review on the development of academic motivation, that this is exactly the issue that needs to be more deeply explored. In their view, "we are in need of putting into practice the policies, interventions, and schooling strategies that emanate from research" (p. 191). They contended that research on academic motivation should place a greater emphasis on ecologically grounded investigations, and less on decontextualized research efforts that

* This chapter is based on: Tempelaar, D. T., Gijssels, W. H., Schim van der Loeff, S., & Nijhuis, J. F. H. (2007). A structural equation model analyzing the relationship of student achievement motivations and personality factors in a range of academic subject-matter areas. *Contemporary Educational Psychology*, 32 (1), 105-131.

ignore differences in students' aptitude, interest, confidence, and classroom processes.

It is remarkable that so little research has been done to investigate the question of how students' motivations are formed and whether motivations are indeed different for different subject domains. This is especially remarkable since research findings indicate that expectancies and values seem to predict career choice (Wigfield & Eccles, 1992). The prime aim of this article is to further our knowledge about students' achievement motivations. More specifically, our research addresses the level of specificity for conceptualizing domains. According to Pintrich and Schunk (2002), this has become a prominent research question in educational research: "Specificity of domains is one of the biggest issues that needs to be resolved for any cognitive or motivational theory that proposes domain specificity of constructs" (p. 64). They pointed out that much research has been done on the role of achievement motivation in learning, but what is severely lacking is research on the level of specificity for conceptualizing domains.

1.1 Models of achievement motivation

The prototype models of achievement motivation are laid out in Atkinson's "Theory of Achievement Motivation" and Lewin's "Resultant Valence Theory" (see Pintrich & Schunk, 2002; Weiner, 1992). In these models performance choices are described as the outcome of a mathematical optimization problem by rational human beings. Atkinson's model, for example, regards behavior as a multiplicative function of the three components: motives, probability of success, and incentive value. All of these components describe personal characteristics. Hence, people come to different performance choices not because the mechanism of choosing is different, but because of individual differences in motives, perceived probability of success, and incentive value. Atkinson's and Lewin's prototype models developed into the current expectancy-value models for achievement motivation in classroom settings (Pintrich & Schunk, 2002). Expectancy-value models take their name from the key role of two components in the motivation to perform an achievement task: students' expectancies for success, and the task value, that is the value they attribute to succeeding the task. These are basically two of the three components in Atkinson's model, leaving out the affective motives component. In current expectancy-value models (Pintrich & Schunk, 2002; Wigfield & Eccles,

2000, 2002), the expectancy component consists of two subcomponents: students' self-perceptions of ability or competence, and students' judgments of the difficulty of the task. Empirical work based on the expectancy-value model, investigating students' motivations for a wide range of subjects (e.g., mathematics, reading, sports, or music) has found that competence beliefs and achievement values are clearly distinct constructs (e.g., Wigfield & Eccles, 2000).

1.2 Schau's (2003) expectancy-value model

The expectancy-value model has met with some criticism (e.g., Pintrich and Schunk, 2002) because it lacks affective constructs in comparison with the achievement motivation model developed by Atkinson. In the present study, achievement motivations are operationalized by adopting a version of the expectancy-value model that incorporates an affective construct, developed by Schau and co-authors (Dauphinee, Schau, & Stevens, 1997; Hilton, Schau, & Olsen, 2004; Schau, 2003; Schau, Stevens, Dauphinee, & DelVecchio, 1995; Sorge & Schau, 2002). Schau's expectancy-value model contains four constructs. The first two expectancy factors deal with students' beliefs about their own ability and perceived task difficulty: Cognitive Competence and Difficulty. The third construct is the subjective task-value: Value. Together, these three constructs constitute the traditional expectancy-value model (Pintrich & Schunk, 2002; Wigfield & Eccles, 2000, 2002). In addition to task-value, Schau and co-authors introduced a second task-related attitude: Affect. Schau's four-factor model of achievement motivations thus strongly resembles early operationalizations of expectancy-value theory, such as Atkinson's model. The re-introduction of the affective component is motivated by the vast evidence of the role of affective factors in learning in mathematics-related domains. Statistics being the single mathematical subject in many social science studies, it is hypothesized that task-related affective attitudes toward statistics are an important component of achievement motivations (Schau, 2003). Empirical research within the domain of statistics (Dauphinee et al., 1997; Hilton et al., 2004; Schau et al., 1995) supports the distinction of Affect and Value. However, it has not been investigated whether Schau's version of the expectancy-value model is valid in the business education domain. This question will be addressed in this study.

The empirical research in the general expectancy-value model referred to above (Wigfield & Eccles, 2000) has typically found that competence beliefs and achievement values differ over domains. Most of this research,

however, focuses on children and adolescents in middle and high school and their motivations for a highly varied range of subjects like mathematics, reading, sports, or music. The range of subjects that are part of a university curriculum is less wide. This may offset the effect of age generally found in empirical research: domain specificity tends to increase with age, which means that competence beliefs and values become more differentiated in time (Wigfield & Eccles, 2000, 2002). This article breaks new ground by researching domain specificity of achievement motivations of university students.

1.3 Domain specificity: Subject-matter areas in business programs

The overall purpose of this article is to investigate the variation in achievement motivations for a range of subjects that are part of a business curriculum. Business programs offer an attractive field of application for research in the variation of achievement motivations since such programs are grounded in three scientific disciplines: mathematics, economics, and behavioral science (Kolb & Kolb, 2005). Students learn about statistics, the most prominent mathematical subject in a business program. Additionally, students also complete coursework in economics such as finance and accounting as well as three subjects grounded in behavioral science. Through this choice, a full coverage of potential domain structures is achieved, from well-structured domains as mathematics and economics to ill-structured domains as the behavioral sciences (e.g., Lawless & Kulikowich, 2006). In this context, we investigate the hierarchical and multi-dimensional nature of achievement motivations. Specifically, we wanted to determine the extent to which different subjects share achievement motivations (i.e., the hierarchical part), and the extent to which achievement motivations are subject-specific, and in this way give rise to multi-dimensionality. From a methodological point of view, the decomposition of achievement motivations into these generic and subject-specific components requires the estimation of second order confirmatory factor models, where second order factors represent generic motivations, and first order factors represent subject-specific motivations.

Domain specificity of achievement motivations is also investigated at the level of categories of subjects in this article. Categories are defined using a reduced Biglan taxonomy of hard and soft subjects (Burke & Moore, 2003). An issue that will be addressed is whether patterns in achievement

motivations for subjects vary more between categories than within a category. The literature referenced above, even the subset elaborating the Biglan classification scheme, is somewhat inconsistent in using and delineating the concepts, *discipline*, *domain*, and *subject*. Although formal definitions do exist (see, e.g., Alexander, 1992), in this article academic areas are consistently referred to as *subjects*. Hereby the fine nuance that might exist between larger areas as statistics and accounting or finance and smaller sub-areas exemplified by the business subjects is ignored.

1.4 Personality factors

Empirical research into the relationship between personality factors and achievement motivations is limited. Generally, such research focuses on goal orientation behavior. Judge and Ilies (2002) reported positive relations for the personality factors conscientiousness and emotional stability with goal-setting motivation. Day, Radosevich, and Chasteen (2003) investigated the relation between personality factors and goal-orientation. They found learning goal orientation to be positively influenced by conscientiousness, emotional stability and especially intellect, but performance goal orientation to be negatively influenced by the same three personality factors and in addition by extraversion. Vermetten, Lodewijks, and Vermunt (2001), studying goal orientations in the framework of ego, task, and effort orientation, found intellect to be a predictor of task orientation, and conscientiousness and agreeability to be predictors of effort orientation. In a very recent study investigating the relationship between a multi-dimensional self-concept measure, Big 5 personality factors, and course achievements, Marsh, Trautwein, Lüdtke, Köller, & Bau (2006) found conscientiousness to be positively related, and intellect to be negatively related, to math self-concept. A similar focus on personality traits and achievement motivations, but a rather different methodological set-up, is found in the work by Ackerman and co-authors (Ackerman, 2003a, 2003b; Ackerman, Bowen, Beier, & Kanfer, 2001). The aim of their research is to find determinants of both cognitive, affective, and conative type for individual differences in knowledge. Amongst potential determinants of individual knowledge, Ackerman and co-authors consider a range of self-report measures, including measures of self-concept, self-estimates of ability, personality, interests, motivational skills and traits, and activities and experiences. Using exploratory factor analysis, the about 40 different self-report measures are reduced to four or five factors or so-called 'trait complexes': constellations of traits of different type that appear to cluster.

E.g., in Ackerman et al. five trait complexes are distinguished, the first two being 'Science/Math/Technology' and 'Verbal/Intellectual'. The personality trait openness loads on the second trait, as does verbal self-concept. Similarly, the personality trait of emotional control loads on the first trait complex, as does the self-concepts for math, science, and spatial self-concept. In this indirect way, relationships between personality traits and achievement motivations are established, which however do not lend themselves easily to an interpretation in terms of a causal relationship between stable personality traits and context-dependent achievement motivations, or between generic and course-specific achievement motivations. In order to allow for such interpretations, the present study will choose for the methodological focus of confirmatory second-order factor analysis and structural equation modeling.

Following the main-stream approach in personality research, we adopted the Big 5 conceptualization of personality (Digman, 1990; Marsh et al., 2006) that distinguishes five major dimensions of personality. The first dimension is indicated as Agreeableness, and comprises aspects of agreeableness, conformity, docility, and friendliness at one end of the dimension, and indifference to others, and self-centeredness at the other end. The second dimension, Conscientiousness, is to be interpreted as the will to achieve. Dimension three, Extraversion, is often referred to as surgency. It contains extraversion and introversion as the two ends. The fourth dimension, Emotional Stability, expresses level of emotional control at one end. At the other end of the dimension, neuroticism, is used to label this dimension. The fifth dimension, Intellect, has been interpreted as intellect, intelligence, and openness to experience at one end. At the other end, the dimension characterizes thinking that is considered closed to new experiences with less cognitive flexibility.

In summary, the current study aims to address the following research questions:

- Can the affect-extended expectancy-value model be generalized to subjects other than statistics such as those representing business programs?
- Are achievement motivations as identified by the affect-extended expectancy-value model, generic in nature, or do they contain subject domain specific components?
- If achievement motivations are found to contain subject-specific components, what can be concluded from second-order factor

analytic models about the relative importance of the subject-specific parts versus the generic parts?

- Can the subject-specificity of achievement motivations be explained in terms of subject-specific relationships between achievement motivations and personality factors through structural equation modeling?
- Does the classification of subject-matter areas into broad groups such as well-structured versus ill-structured add plausibility to the results in the sense that similar subjects demonstrate similar patterns?

In answering these research questions, the appropriate methodological tools are those of structural equation modeling. The conceptualization of psychological variables as latent constructs of which the measurement necessarily will contain errors, requires the use of statistical methods designed for the latent variable case.

2. METHOD

2.1 Participants

In this longitudinal study, students are followed who participate in the program, International Business, of Maastricht University. In their second year, students take eight required intermediate level courses in core business and auxiliary subjects. In four of these courses, a large majority of students participated in a survey investigation on subject achievement motivations. These were three business courses as well as a course on finance and accounting. Table 5.1 presents a partial outline of the curriculum of this undergraduate program, the timing, and the number of students who participated in the survey investigation. Before entering the second year, these students, together with students of a parallel program International Economics, take required courses in the same subjects at an introductory level, and some more service courses. Our research started in one of these service courses. Specifically, the first year course, Quantitative Methods (QM), is a required introduction into mathematics and statistics for all business and economics students, with approximately 1000 students participating. In that QM course, the very first subject achievement motivations questionnaire is administered assessing achievement motivations for statistics, with 972 respondents. Of those 972 respondents,

574 (59%) reported being male and 393 (40%) reported being female. Other relevant decompositions of the sample are: 522 students (54%) reported having a Dutch secondary school diploma, while 427 students (44%) reported having non-Dutch diplomas (most of them of German nationality).

Table 5-1. Partial outline of undergraduate program, specifying the courses under study, and response on motivation survey, respectively motivation & personality surveys

Period	Course names	SATS response	SATS & Big5 response
Semester 1/2	Quantitative Methods	N=972	N=522
Semester 3	Business Strategy	N=368	N=332
Semester 3	Finance & Accounting	N=372	N=287
Semester 4	Marketing Management	N=291	N=248
Semester 4	Organization & HRM	N=304	N=249

All motivation surveys were administered in the very first session of each of the courses to ensure that responses reflect students’ attitudes toward the academic subject only and are not influenced by impressions of the educational process. All courses are intermediate level courses, so measured achievement motivations are to be regarded as students’ motivations for subjects resulting from learning similar subjects at an introductory level.

2.2 Measures

2.2.1 Subject achievement motivations

Achievement motivations for statistics are measured with the Survey of Attitudes Toward Statistics (SATS) developed by Schau and co-authors (Dauphinee et al., 1997; Hilton et al., 2004; Schau, 2003; Schau et al., 1995; Sorge & Schau, 2002). The SATS consists of 28 seven-point Likert-type items measuring four aspects of post-secondary students’ statistics attitudes. The SATS contains four scales, as indicated below. Each scale is accompanied by two examples of items. One example is worded positively while the other example is worded negatively.

- Affect (six items) measures positive and negative feeling concerning statistics. Examples include: I like statistics, and I am scared by statistics.

- Cognitive Competence (six items) measures attitudes about intellectual knowledge and skills when applied to statistics. Examples include: I can learn statistics, and I have no idea of what's going on in statistics.
- Value (nine items) measures attitudes about the usefulness, relevance, and worth of statistics in personal and professional life. Examples include: I use statistics in my everyday life, and I will have no application for statistics in my profession.
- Difficulty (seven items) measures attitudes about the lack of perceived difficulty of statistics as subject. Examples include: Statistics formulas are easy to understand, and Statistics is highly technical.

Schau et al. (1995), Dauphinee et al. (1997), and Harris and Schau (1999) elaborated on the development process of the instrument. The instrument is freely available from the internet (Schau, Dauphinee, Del Vecchio, & Stevens, 1999, <http://www.unm.edu/~cschau/downloadsats.pdf>). Validation research has shown that a four-factor structure provides a good description of responses to the SATS-instrument in two very large samples of undergraduate students (Dauphinee et al., 1997; Hilton et al., 2004).

Achievement motivations for the other four subjects (i.e., Business Strategy, Finance & Accounting, Marketing Management, and Organization & Human Resource Management (HRM)) are measured by an adapted version of the SATS developed by the authors of this study. The adaptation of the SATS inventory was done by replacing the subject name Statistics by each of the four subject names mentioned. Additionally, two items were slightly reformulated to accommodate the different nature of business subjects. The statement, Statistics formulas are easy to understand, was rephrased as Techniques of Business Strategy are easy to understand. Similarly, the statement, Statistics involves massive computations, was reworded as Business Strategy involves a lot of technical aspects.

2.2.2 Personality

As a personality instrument, the 50-item version of the IPIP Big 5 factor unipolar markers (Goldberg, 2005; Gow, Whiteman, Pattie, & Deary, 2005) was used. The instrument consists of 10 items for each of the Big 5 personality factors: Extraversion, Agreeableness, Conscientiousness, Emotional Stability (the opposite pole of Neuroticism) and Intellect (sometimes labeled as Openness to Experience). The instrument is freely available on the internet (Goldberg, 2005, <http://ipip.ori.org/ipip/>).

2.3 Procedure

2.3.1 Classification of academic subject-matter areas

In order to allow for formulating research hypotheses of differentiated motivations toward several academic subjects, we needed a classification of these subjects. One of the best known classification schemes is the three-dimensional scheme of Biglan: 1) hard-soft; 2) pure-applied; and, 3) life-nonlife system (Biglan, 1973a, 1973b; Schommer-Aikins, Duell, & Barke, 2003). Biglan's first dimension of hard versus soft academic disciplines is based upon the degree to which a clearly delineated paradigm exists. Hard subject matter areas, such as mathematics, have a commonly agreed upon set of problems for study and accepted methods to be used in their exploration. Biglan's second dimension, pure versus applied areas, is based on the extent of the practical application of the subject matter. Biglan's third dimension, life system versus nonlife system areas, refers to the level of involvement with living objects of study. Biglan's model was derived from the statistical analysis of scholars' perceptions of similarities among 35 academic departments in one institute (Biglan, 1973a). His original classification contained among others the subjects areas mathematics (i.e., pure & hard & nonlife), accounting, finance, and economics (i.e., all applied & soft & nonlife), and psychology and sociology (i.e., both pure & soft & life). Subsequent empirical research has extended Biglan's classification. For example, more subject areas have been added to the original 35. In the context of this study, the adoption of business subjects as soft, applied and life subject-matter areas is especially relevant (Burke & Moore, 2003; Paulsen & Wells, 1998; Schommer-Aikins et al., 2003).

Due to the choice of curriculum subjects in this study, and the strong differences between the grounding scientific disciplines, we can reduce the three-dimensional scheme to a one-dimensional ordering (e.g., Burke & Moore, 2003). Specifically, we placed hard, pure, and nonlife as one pole of the spectrum versus soft, applied and life as the other pole. This dimension-reduced taxonomy strongly resembles the one-dimensional classification used by Kulikowich and DeFranco (2003) (i.e., well-structured, rule-based domains to ill-structured). Alexander (1992, 1997) used similar descriptions in her investigations on complex domains.

In this scheme, the three business subjects would be considered ill-structured and complex. In contrast, statistics, as a domain, can be characterized as well-structured, rule-based along with the subject-matter areas of accounting and finance.

2.4 Data analysis

2.4.1 Missing data and data availability

Data used in this study were based on self-reports. For these data, two different types of missing data can be distinguished. The first type of missing data stems from the students who do not answer all items of the questionnaire. If at most one item for each parcel (see description in the next section) is missing, then missing data were replaced by the parcel mean (the parcel score was calculated excluding the missing item). In total, only .14% of all items fell in this category. If more than one item per parcel were missing, all data belonging to the relevant self-report were deleted. This first step of handling missing data was applied for all models tested.

The second source of missing data was where a student missed the first session of any of the five courses, and thus missed the administration of the subject motivation list. The number of students who participated in all five subject motivation administrations was 153, with another 111 students missing exactly one survey. To increase data availability for the single-sample models estimating achievement motivations simultaneously for all five academic subject-matter areas, we applied the PRELIS Multiple Imputation procedure on the complete data matrix of parcels to impute values for the single missing survey of 111 students. This is a model-based imputation method based on the expectation-maximization (EM) algorithm. In this technique, imputed values are chosen to best match the pattern of other cases over a set of matching variables (Kline, 2005; Schumacker & Lomax, 2004). The data set resulting from this imputation procedure, ($n = 264$), will be referred to as the final data set.

2.4.2 Parceling

As a first step, negatively worded items were reverse-scored, such that for all items a higher score corresponded to a more positive achievement motivation or personality trait. This is also true for the items in the motivation scale Difficulty, making the name of this scale somewhat

confusing. Perceived lack of difficulty would have been more appropriate. As a subsequent step, items from the self-report instruments were parceled. The technique of item parceling, where items from the same subscale are aggregated into several parcels or mini-scales, has been adopted in empirical studies for several reasons including: a) obtaining more continuous and normally distributed observed data; b) reducing the number of model parameters to achieve a more attractive variable to sample size ratio; and, c) estimating stable parameters (Bandalos, 2002; Coffman & MacCallum, 2005; Hau & Marsh, 2004; Marsh, Hau, Balla, & Grayson, 1998). In our study, the size of the model relative to the sample size necessitates in itself the parceling step in the estimation of the second-order factor model of subject-matter area motivations.

In parceling items, Hau and Marsh (2004) advised to counterbalance skewness in the case of strong non-normality by creating parcels out of item pairs with opposite skew. As a preliminary step to parceling, the degree of non-normality of the data was determined. In the motivations data, no indications of non-normality were found in any of the self-report questionnaires beyond Hau and Marsh's category of moderately non-normal (i.e., skew = 1.0 and kurtosis = 1.5). Non-normality in Big 5 data was at even lower levels. In addition, all scales satisfied the typical pattern of self-report data described by Hau and Marsh of being slightly negatively skewed, thereby making the counterbalancing scheme of parcel construction hardly relevant.

In the empirical analyses of their SATS data, Schau et al. (1995), Dauphinee et al. (1997), and Hilton et al. (2004) adopted item parceling based on balancing with respect to the positively and negatively worded items, size of parcel means, standard deviations, and skew (see Schau et al., 1995). Because non-normality in our data does not provide strong arguments to adopt a parceling scheme based on counterbalancing skew, we parceled our motivation data in exactly the same way as proposed in Schau et al. (1995). The parceling scheme for the Big 5 data was based on the same mechanism of balancing positively and negatively worded items. Because items included in the Big 5 are in alternating order in terms of their wording, it was decided to parcel the 10 items per subscale into five parcels each consisting of a pair of consecutive items.

2.4.3 Goodness of fit

With large sample sizes as in our study, the χ^2 test statistic is prone to model rejection in virtually any formal test of significance (Byrne, 1998; Marsh & Yeung, 1996). Following the seminal work on cutoff criteria for fit indexes of Hu and Bentler (1999), emphasis is placed on the root mean square error of approximation (RMSEA), the standardized root mean squared residual (SRMR), and the Comparative Fit Index (CFI). The choice of a cutoff value close to .95 for CFI, a cutoff value close to .08 for SRMR, and a cutoff value close to .06 for RMSEA, appears to provide best guidelines for establishing good model-data fit in the simulation studies performed by Hu and Bentler. The combination of these three cutoff criteria was therefore used to determine model-data fit in this investigation. Since in empirical research in achievement motivations other researchers have used additional fit indexes (e.g., Marsh & Yeung, 1996; Hilton et al., 2004), three more indices will be reported: 1) the Non-Normed Fit Index (NNFI; termed Tucker-Lewis Index or TLI in Marsh & Yeung, 1996); 2) the Goodness-of-Fit Index (GFI); and, 3) the Relative Fit Index (RFI, termed Relative Noncentrality Index or RNI in Marsh & Yeung, 1996).

2.4.4 Statistical analyses

This study integrates several techniques of structural equation modeling (SEM). A structural equation model is distinct from a measured variable path model in that it hypothesizes crucial variables (e.g., achievement motivations), may not be directly observable and are better modeled as latent variables. SEM makes it possible to distinguish two different types of errors: errors of measurement in the observation of variables, and errors of prediction in structural equations. Models that involve only observed variables, as regression and path models, assume that measured variables are perfectly valid and reliable. In applications like this one, where all measurements take place through self-report questionnaires, the assumption of perfect validity and reliability of scores is clearly unrealistic. Measured scales of our psychological constructs contain a measurement error component of an unknown size that is likely to vary for different instruments. If measurement error is present but not accounted for in the model estimation step, this might have serious consequences (Schumacker & Lomax, 2004). Therefore, all constructs under study are assumed to contain measurement errors, making the SEM approach appropriate.

Models were estimated with LISREL (version 8.72) using maximum likelihood estimation.

Models used in this study either describe characteristics of a single subject domain, or simultaneously describe all academic subject-matter areas. From a statistical point of view, the estimation of subject domain specific models is rather different from the estimation of models describing all subject domains simultaneously. Subject domain specific models are small, and are estimated on relatively large data sets. In contrast, models for several subject domains are complex, and are estimated on relatively small, composite data sets. Therefore, empirical outcomes are subdivided over five results sub-sections. The 1st, 2nd, and last describe subject-matter area specific models while the 3rd and 4th simultaneously model all subject-matter areas.

In order to investigate whether the affect-extended expectancy-value model can be generalized, the first results section starts with a descriptive analysis of achievement motivations in the various academic subjects. Subsequently, we examine how well a four-factor model of achievement motivations, validated for the subject statistics in several studies, describes achievement motivations for several academic subjects. We do so by applying first-order confirmatory factor analysis (CFA) to data sets corresponding to each of the subjects. More specifically, the type of model we apply is that of a Correlated Trait (CT) first-order CFA. In this analysis, latent factors or traits are allowed to covary, and are therefore not restricted to be orthogonal. The correlation of latent traits is preeminently an interesting characteristic of the subject-matter areas, since it indicates the existence of commonalities in achievement motivations for different subjects. The aim of this phase is not only to ascertain that the four factor structure for subject achievement motivations explains the data well, but also to provide a so-called baseline model for the next phase of the analysis: analyses to discover similarities and differences among academic subject-matter areas (Byrne, 1998). A rather strict interpretation of the baseline model was opted, for the baseline model should provide good fit in all five subjects but still be parsimonious. Parsimony is interpreted as containing no single nonsignificant parameter estimate in any of the subject-specific models. Once the baseline model is determined, it is estimated on data collected for each of the academic subjects separately. In the estimation of these subject-specific measurement models, ratios expressing the amount of data available relative to the size of the model are favorable. The number of cases (sample size) ranges from 291 to 972 (see

Table 5.1, third column). The number of free parameters is 23. By result, the ratio of cases to free parameters ranges from 13 to 42, above the commonly advised five cases per free parameter (e.g., Bentler & Chou, 1987). Further, given that the baseline model has only four factors, the ratio of cases to factors ranges between 73 and 243, far exceeding the small sample sizes often allowable in CFA (see, e.g., Marsh et al., 1998).

Eyeball comparison of the five measurement models indicates that achievement motivations are not invariant over academic subjects. Formal tests on multi-sample invariance (Kline, 2005; Schumacker & Lomax, 2004) or multi-group invariance as described the work of Byrne (1998), requires the assumption of independent groups of observations. In our study, this assumption is implausible, since observations in the various subjects are based on the same group of students. Therefore, the analysis of similarity and dissimilarity of achievement motivations in various academic subject-matter areas is continued in the context of single-sample or one-group SEM. In the single-sample framework we explicitly take into account that the observations stem from the same group of students.

2.5 First-order CFA model

Two different types of measurement models for the simultaneous estimation of all subject-specific achievement motivations are used to investigate commonalities in motivations, namely a first and second-order CFA. The first type of model basically collects all five subject-specific first-order CFAs for achievement motivations into one large first-order CFA. This first-order CFA is again of the CT type, be it that the structure imposed on the trait correlations is more complex than in the single subject models. The first form of trait correlation allowed is the one derived from the baseline model for single subject models. All pairs of achievement motivations within individual subjects are allowed to correlate, except for the Value, Difficulty pair, which was restricted to zero because of nonsignificant estimates in some subjects. The second form of trait correlation tested is that corresponding to the same achievement motivations for different subjects. Beyond trait correlations of different motivations for the same subject and trait correlations of the same motivations for different subjects, all other entries of the trait correlation matrix (i.e., the correlations between different achievement motivations for different academic subjects) are a priori restricted to zero. The prime reason to do so relates to the issue of data availability and the need for parsimonious modeling. The number of cases available for the estimation of the CT, first-order CFA model in the single-

sample framework is 264. However, the number of factors rises from 4 in a single-subject model to 20 (i.e., 5×4) in the simultaneous subject model, and the number of free parameters from 23 to 148, even with the above mentioned restrictions on the trait correlation matrix. As a consequence, the ratio of cases to factors becomes 12.3, still well above the advised range of 5-10 in exploratory factor analysis (see, e.g., Marsh et al., 1998). The ratio of cases to free parameters is 1.8, which is less than the frequently used rule of thumb of “several cases per free parameter” (Marsh et al., 1998). The totally unrestricted CT model contains more free parameters than the number of available cases. Recent simulation research on sample size recommendations in CFA suggests that the number of free parameters being higher than the number of cases is in itself not problematic (Gagné & Hancock, 2006; Marsh et al., 1998). More important than ratios as cases to factors or cases to free parameters, is the consideration of model quality. Construct reliability is a main feature of model quality, and Gagné and Hancock have developed benchmark values for minimal construct reliability for first-order CFA models. However, model structures analyzed within the single-sample framework of our study are more complex than the structures analyzed in Gagné and Hancock’s simulation study. Therefore, we opted for the somewhat conservative choice of a restricted CT structure for our baseline model in the single-sample framework. A further advantage of adopting a restricted CT structure is that model estimation is strictly confirmatory and framed within factor and parameter structures already corroborated as the baseline model for individual academic subjects. A further and even more parsimonious modeling option is provided by a so-called ‘CTCM’ or correlated trait, correlated method structure, based on a multi-trait, multi-method model (e.g., Byrne, 1998; Kline, 2005; Schumacker & Lomax, 2004). This last model is more parsimonious since it distinguishes only nine latent factors (i.e., the 4 achievement motivations, and the 5 academic subject-matter areas). In this model structure, every subject-specific achievement motivation is hypothesized to be linearly composed of a generic achievement motivation component, and a subject component. Measurement models of motivations based on the CTCM structure appeared to demonstrate unsatisfactory fit, and will not be elaborated for that reason in this study.

2.6 Second-order CFA model

The second type of model investigated in the single-sample framework is the CT, second-order CFA model. This model is again estimated on the data set of students having provided responses on all motivation surveys (i.e., $n = 264$). Thus, similar data-related concerns are in place as in the estimation of the first-order model. Compared to the first-order model, the number of factors in the second-order model grows to 24 (i.e., added to the 20 first-order factors are four second-order factors). The number of free parameters however diminishes. That is so because most correlation terms introduced in the first-order model to represent the relationship between like motivations for different subjects are now imposed by only five correlation terms between second-order factors of motivations. Estimating this second-order CFA allows analysis of commonalities in motivations for subject-matter areas beyond a pairwise level. For all subject-matter areas under study, variation in motivations is decomposed into a subject-specific component as well as a generic component generated by the second-order factors. We use the word, generic, to indicate commonalities among all subjects, whereas common will refer to pairwise analyses.

In the last results section we return to the specification of subject-specific models, where the full structural model is estimated with personality factors explaining achievement motivations. In contrast to the analyses reported in the earlier results sections, the modeling reported in this section contains a model search step, an exploratory element. Measurement parts for both motivations and personality factors used in the full model are still confirmatory in nature. Both are a CT, first-order CFA. In agreement with the two-step model building approach widely advocated (e.g., Schumacker & Lomax, 2004), the measurement models are estimated prior to the estimation of the structural model. The structural part of the full structural equation model is not a priori restricted, except for the hypothesis with regard to the direction of the relationship. Specifically, stable personality traits are hypothesized to be causal for context dependent achievement motivations. For the estimation of these structural parts, two different model modification procedures are applied. The first is called model trimming by Kline (2005) and backward search by Schumacker and Lomax (2004). Starting from a full matrix of structural path coefficients, one by one, parameters are restricted to zero if they prove non-significant, until all remaining structural parameters are significant. The second approach is called model building by Kline (2005) and forward search by Schumacker and Lomax (2004). It starts from a zero matrix of structural paths

coefficients, and frees parameters one by one, in the order indicated by the value of the modification indices, up to a point where no more significant improvement in fit is achieved.

Because in both approaches subsequent models are nested, the χ^2 difference statistic can be used to assess model fit. For all five subject-matter areas, both forward and backward searches converge to the same final model. Model modification is a form of exploratory analysis, and may impact Type II error rates regarding the acceptability of models as a whole. Effects on such Type II error rates is stronger the smaller the sample size (Kline, 2005). However, because the full structural equation models are estimated on individual subject-matter area data, relevant sample sizes are those reported in the last column of Table 5.1. The resulting cases to factors ratios range from 29 to more than 80, the cases to free parameters ratios from 3.2 to 6.5.

3. RESULTS AND DISCUSSION

3.1 Psychometric properties of achievement motivation scores

The SATS instrument (Schau et al., 1995) measures achievement motivations for statistics in the context of the expectancy-value theory. The SATS has been adapted for the subjects of statistics by Schau and coauthors. An affective construct was introduced in the instrument in addition to the expectancy construct, the perception of competence construct, and the task value belief construct (i.e., the constructs familiarly used in expectancy-value models). In empirical applications reported in Schau et al. (1995), Dauphinee et al. (1997), and Hilton et al. (2004), it has been confirmed that making the distinction between affect and value is important in explaining the variation in achievement motivations.

The descriptive statistics in Table 5.2 show that academic subject-matter areas considered hard according to the Biglan classification, namely Statistics and Finance & Accounting, have lower achievement motivations than the other subjects. In addition, the variation in Affect for both hard subject-matter areas is larger than the variation in Affect for the soft subject-matter areas, indicating greater diversity. A second descriptive observation

is that for all subject-matter areas Value is the highest scoring achievement motivation, and (lack of perceived) Difficulty the lowest.

Table 5-2. Scale means and standard deviations for achievement motivations of students in final data set

	Affect		Cognitive Competence		Value		Difficulty	
	M	SD	M	SD	M	SD	M	SD
Statistics	4.42	0.88	4.76	0.75	5.14	0.70	3.45	0.60
Business Strategy	5.48	0.75	5.61	0.67	5.57	0.69	4.33	0.70
Finance & Accounting	4.01	1.11	4.46	0.93	4.84	0.86	2.78	0.63
Marketing Management	5.45	0.68	5.41	0.70	5.49	0.69	4.23	0.59
Organization & HRM	5.34	0.66	5.32	0.66	5.21	0.73	4.59	0.63

A check on the reasonableness of composing achievement motivation surveys for subject-matter areas other than statistics is to compare the Cronbach coefficient α values for the several subject / scale combinations with values reported in other studies. Table 5.3 presents all reliability indices; reference α values obtained from Schau et al. (1995), Dauphinee et al. (1997), and Hilton et al. (2004) are summarized into one row indicating the spread of values. The alpha coefficients found in our study are, although somewhat lower than the reference values, satisfactory compared to general findings reported elsewhere in the literature. The only exception is the α for Difficulty with regard to Statistics, being 0.62 in value. It should be noted that in some studies the reference values were obtained by deleting extreme cases from the data set. This has not been done in this study.

Table 5-3. Cronbach Coefficient α Values for achievement motivation scales in different subjects, and reference values from 11 studies as reported in Schau (2003)

	Affect	Cognitive Competence	Value	Difficulty
Statistics	.80	.72	.76	.62
Business Strategy	.71	.76	.79	.68
Finance & Accounting	.82	.80	.83	.72
Marketing Management	.77	.79	.83	.68
Organization & HRM	.72	.73	.82	.73
Reference values	.80-.89	.77-.88	.74-.90	.64-.81

The most remarkable outcome is that, except for the Affect scale, the four business subject-matter areas seem to have higher reliabilities than Statistics itself. Judged from the criterion of reliability, it seems that measuring motivations, based on the adapted SATS, provide coherent scales. As a next step, we will investigate how well a four-factor model explains the variation in motivations in each of the subjects.

3.2 Measurement models for achievement motivations

The first research question mentioned in the introductory section addresses the generalization of the four-factor model developed for the affect-extended expectancy value structure to academic subjects other than statistics. In order to answer that research question, confirmatory factor analysis allowing correlated traits (CT) was applied to all subject achievement motivations data. The five academic subject-matter areas are described in Table 5.1, with the third column indicating relevant sample sizes of the subject-matter models. Table 5.4 contains fit indices of the baseline version of these CT, first-order factor models, and Table 5.5 the correlation structure of latent motivational factors.

Table 5-4. Fit indices of four-factor confirmatory factor models of achievement motivation for five academic subjects

	χ^2	df	RMSEA	SRMR	GFI	NNFI	CFI	RFI
Statistics	63.97	22	.044	.049	.99	.99	.99	.98
Business Strategy	35.04	22	.041	.036	.98	.99	.99	.97
Finance & Accounting	43.59	22	.050	.061	.98	.99	.99	.98
Marketing Management	65.13	22	.083	.036	.95	.96	.97	.94
Organization & HRM	59.47	22	.079	.037	.96	.97	.98	.95

Table 5.5 clearly shows that 5 out of 6 trait correlations are significantly positive for all subject-matter areas. Only one trait correlation, (i.e., Value and Difficulty), is significant in only 2 of the 5 subject-matter areas. For this reason, in the baseline model this sixth trait correlation is restricted to zero, while the other five correlations were set free, in agreement with the requirements of the baseline model. The resulting four-factor model of achievement motivations for business subjects fits the data very well. The expectancy-value theory on which the adapted SATS based achievement motivations models are designed clearly has broader validity than for the

statistical domain. Across the domains, fit indices in Table 5.4 demonstrate very little variability.

Table 5-5. Estimated latent factor correlations of achievement motivations for five business subjects

	Affect	Cognitive Competence	Value	Difficulty
Statistics				
Affect	1.00			
Cognitive Competence	0.91***	1.00		
Value	0.42***	0.50***	1.00	
Difficulty	0.67***	0.71***	0.14**	1.00
Business Strategy				
Affect	1.00			
Cognitive Competence	0.81***	1.00		
Value	0.69***	0.64***	1.00	
Difficulty	0.40***	0.59***	0.07	1.00
Finance & Accounting				
Affect	1.00			
Cognitive Competence	0.92***	1.00		
Value	0.54***	0.58***	1.00	
Difficulty	0.63***	0.67***	0.15*	1.00
Marketing Management				
Affect	1.00			
Cognitive Competence	0.81***	1.00		
Value	0.60***	0.57***	1.00	
Difficulty	0.35***	0.53***	0.03	1.00
Organization & HRM				
Affect	1.00			
Cognitive Competence	0.83***	1.00		
Value	0.74***	0.48***	1.00	
Difficulty	0.39***	0.72***	0.01	1.00

Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

Having gathered support for all four-factor models, the correlation structure of the latent factors depicted in Table 5.5 deserves prime interest. The first issue that comes up is what the effect is of disentangling of the broad task

value concept into two separate concepts: Affect, related to liking the subject, and Value, related to the importance attached to the subject. For all subject-matter areas, the correlation between latent factors Affect and Value is, relative to other correlations, modest. In no instance is this the highest correlation. This indicates that Affect and Value are clearly empirically distinguishable constructs. In these correlations, an interesting pattern with regard to the hard/soft dimension can be observed. The correlation is much higher for soft (average value 0.68) than for hard subjects (average value 0.48). Apparently, disentangling Affect and Value is especially relevant for hard subjects. For these subjects, students can find a subject important but at the same time may or may not like the subject matter. For soft subjects, liking a subject and the attached importance seem to go more strongly hand in hand.

For any given subject-matter area, the correlation between Value and (lack of perceived) Difficulty is the weakest, indicating that in all subject-matter areas the attached value is relatively independent of the lack of perceived difficulty. For the soft subject-matter areas, the correlation is very low. Value and Difficulty are independent in the statistical sense. For the hard subject-matter areas, there is still, be it very modest, a relationship between the two constructs. Those subjects are somewhat better valued by students who regard the subject-matter as attainable.

A third observation refers the correlation between Affect and Cognitive Competence. For all subject-matter areas, this is by far the largest correlation. This is an important finding. Affect is achieved by decomposing the task value component into affective and utility-related factors, but from this analysis it appears that Affect is much more strongly related to the expectancy component Cognitive Competence than to Value. This once again confirms the usefulness of the affect extension of the expectancy-value model. In contrasting hard and soft subject-matter areas, it is evident that the correlation is much stronger in hard subject-matter areas than in soft. For hard subjects, we found an average correlation of 0.92. This is comparable to the results found in Dauphinee et al. (1997) and Hilton et al. (2004) in the statistics domain. In fact, the correlation they found is even somewhat higher, and led them to discuss the issue of whether a 3-factor model, achieved by collapsing the factors Affect and Cognitive Competence into one factor, would fit the data equally well as a 4-factor model. That turns out not to be the case, neither in their studies nor in our investigation. For soft subject-matter areas, it is apparently possible to be confident in one's performance without liking the subject.

3.3 A correlated traits first-order factor model for achievement motivations for academic subject-matter areas

Having established the appropriateness of the four-factor model for achievement motivations in a range of subjects, we are ready to proceed with the second research question: what is the nature of achievement motivations? The derivation of subject-specific measurement models in the last subsection strongly suggests that achievement motivations are not invariant over academic specializations. This unequivocally shows that achievement motivations contain specific components per subject-matter area. It does, however, leave open the possibility that achievement motivations of different content areas share a common component. In order to ascertain whether common components are present, as formulated in our second research question, a 20-factor model is estimated where factors are allowed to correlate according to a specific pattern, using the final data set of 264 cases. Previously, we have indicated this model as a CT, first-order CFA. That is, traits correlate if either they correspond to the same academic subject, or to the same motivation. All other trait correlations are a priori restricted to zero. With regard to the trait correlations within subject-matter areas, the structure of the baseline model was imposed. That is, five correlations were set free, the sixth was restricted to zero. Since our study comprises data on five different subjects, the number of subject pairs to be investigated equals 10. For each pair, four traits are allowed to correlate. Of the 4 x 10 trait correlations, 35 proved to be significant at a .05 level. Table 5.6 provides fit indices of the CT first-order CFA. Fit indices provide a mixed picture but certainly leave room for an improvement in modeling the structure of trait correlations.

Table 5-6. Fit indices of correlated trait, first-order factor model of achievement motivations for five academic subjects

	χ^2	df	RMSEA	SRMR	GFI	NNFI	CFI	RFI
CT, first-order CFA	1689.92	880	.060	.13	.78	.93	.94	.87

Table 5.7 contains the estimated trait correlations corresponding to the correlated traits among subject-matter areas. The correlations depicted in Table 5.7 represent the commonalities of corresponding motivations for 10 different pairs of academic subjects. There are strong differences among the four motivations with regard to the size of the trait correlations. Value

has by far the strongest commonality. The average of all 10 pairwise trait correlations was .32. The average trait correlations of Cognitive Competence and Difficulty were .18, and .19, respectively, whereas the average trait correlations of Affect were .08 with 5 out of 10 correlations being nonsignificant.

Table 5-7. Same subject trait correlations matrix of CT, first-order CFA

	Statistics	Business Strategy	Finance & Accounting	Marketing Management	Organization & HRM
Affect					
Statistics	1.0				
Business Strategy	.12**	1.0			
Finance & Accounting	.07*	.03	1.0		
Marketing Management	.10*	.22***	.04	1.0	
Organization & HRM	.04	.06	.03	.07	1.0
Cognitive competence					
Statistics	1.0				
Business Strategy	.25***	1.0			
Finance & Accounting	.10*	.15***	1.0		
Marketing Management	.14***	.21***	.14***	1.0	
Organization & HRM	.18***	.21***	.13***	.25***	1.0
Value					
Statistics	1.0				
Business Strategy	.37***	1.0			
Finance & Accounting	.44***	.30***	1.0		
Marketing Management	.27***	.47***	.24***	1.0	
Organization & HRM	.14***	.32***	.19***	.41***	1.0
Difficulty					
Statistics	1.0				
Business Strategy	.28***	1.0			
Finance & Accounting	.23***	.17***	1.0		
Marketing Management	.16***	.26***	.13**	1.0	
Organization & HRM	.18***	.21***	.05	.24***	1.0

Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

Value has by far the largest common component, whereas the common component in Affect was very small and nonsignificant in most cases. This is remarkable given their shared background. Apparently, the isolation of Affect from a broader task value construct has resulted in Value picking up most common components, while Affect represents the subject-specific components. Distinguishing Affect from Value as two separate task value constructs in the expectancy-value model was motivated by research in statistics education. Our data suggest that this decomposition is also relevant outside the statistics domain.

Having established that motivations contain a common and a subject-specific component, an interesting question is how strong the common component really is. One way to investigate this is to compare trait correlations between subject-matter areas with trait correlations within academic specializations. Specifically, in order to address this question entails comparing and contrasting Table 5.7 with Table 5.5. It appears that trait correlations are larger within subject-matter areas than between subject-matter areas. An exception must be made for some correlations within Value. So, although achievement motivations constitute a multidimensional concept, and this study contributes to enlarging this dimension even further, the variability in achievement motivations over these different dimensions is still small compared to the variability in achievement motivations over different subject domains.

The pattern of trait correlations represents our hard/soft classification of academic subjects rather well. The courses Business Strategy and Marketing Management have relatively high commonalities in all four achievement motivations. The pairs Marketing Management and Organization & HRM as well as Business Strategy and Organization & HRM exhibit these commonalities in three out of four motivations. The two subjects classified as hard, Statistics and Finance & Accounting, primarily share the motivation Value.

3.4 Second-order factor model for achievement motivations for academic subjects

The existence of correlated traits suggests that motivations for academic subjects share a generic component. The common components found in the last section, however, are the result of pairwise comparison of subjects. Our third research question is whether these shared components are

generic (i.e., shared by all subject-matter areas). The model that results from this hypothesis imposes restrictions on the correlation structure of the latent factors that was found in the last section. As a last step in the specification search, we estimated such a model – a second-order factor model with a generic second-order factor for each of the achievement motivations – explaining the corresponding subject-specific first-order motivations factors. The aim of this last step is to decompose motivations for academic subjects into a generic component and a subject-specific component, thereby answering the third research question about the relative size of both components.

Initially, in the estimation of the model, all second-order trait correlations were set free. As in the trait correlation structures of the baseline first-order model, all factor correlations appeared to be significant, except for the correlation between Value and Difficulty. Restricting that correlation term to zero, two versions of a second-order CFA were estimated on the final data set of 264 cases. In the pure second-order CFA, the residual correlations of the 35 trait relationships between first-order traits, that were introduced in the CT, were restricted to zero (e.g., Byrne, 1998). This implies that correlation of first-order traits is fully explained by the imposed effect of second-order trait correlations and the relations between first- and second-order traits. These imposed trait correlations provide a good approximation for 31 out of 35 first-order trait correlations. Only for four pairs of traits, complementary correlation terms for latent residuals in first-order traits were required to adequately model covariation of traits at the first-order level. The model labeled as the final second-order CFA therefore contains 5 second-order traits correlation terms and 4 added correlation terms for first-order trait residuals (beyond the within subject first-order factor residuals inherited from the subject-specific factor models). Table 5.8 contains fit indices of these two second-order factor models.

Table 5-8. Fit indices of second-order factor model of achievement motivations for five academic subjects

	χ^2	df	RMSEA	SRMR	GFI	NNFI	CFI	RFI
Pure second-order CFA	1594.19	895	.053	.074	.79	.94	.95	.88
Final second-order CFA	1543.13	891	.051	.072	.80	.95	.95	.89

The fit indices of the final second-order CFA are acceptable. Because the final second-order CFA model is not nested in the CT, first-order CFA model discussed previously, the restrictions imposed by the second-order

model cannot be formally tested. However, on the basis of the fit statistics in Table 5.8, which show an important improvement over those in Table 5.6, we accept the second-order factor model as an adequate description of the data. The final and pure second-order CFA models, however, are nested, so we can formally test the improvement in fit resulting from the introduction of four additional first-order trait residual correlation terms by means of a χ^2 test. That improvement is clearly significant (i.e., $\Delta\chi^2 = 51$ for $df = 4$). Table 5.9 exhibits the estimated correlations of second-order traits. Figure 5.1 depicts the estimated model graphically, with the four second-order latent factors on the left hand side. The 20 subject-specific first-order factors are presented in the middle of the graph, and the 45 subject-specific item-parcels used as indicator variables are on the right-hand side.

Table 5-9. Estimated second-order trait correlations of second-order factor model

	Affect	Cognitive Competence	Value	Difficulty
Affect	1.00			
Cognitive Competence	.79***	1.00		
Value	.57***	.55***	1.00	
Difficulty	.45***	.57***	.05	1.00

Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

The complex structure of the second-order CFA, together with the relatively small sample size in the single-sample framework, result in ratios for model complexity that are modest in size. However, as indicated in the Method section, recent research has shifted away from these traditional model complexity ratios as cases per variable or per parameter, toward consideration of model quality. For example, Marsh et al. (1998) proposed the reliability coefficient Omega for a set of indicators as a measure of construct reliability. Similarly, Gagné and Hancock (2006) suggested an alternative measure, the coefficient H or maximal reliability as another alternative for evaluation of model quality. For the second-order factor model, both these measures were calculated for both first and second-order latent factors. The average value of coefficients Omega and H for the 20 first-order factors are 0.79 and 0.80, respectively. These values are in line with the minimal benchmark values for satisfactory convergence, provided in Gagné and Hancock (2006) for the case of sample size 200 and three indicators per factor (i.e., .71 and .79). The average value of coefficients Omega and H for the four second-order factors were .43 and .76,

respectively. Again, this is in line with the benchmark values, given the larger number of five indicators per factor for the second-order factors (i.e., .43 and .72). These comparisons suggest adequate model quality, be it that benchmark values are developed for the first-order CFA case, while in this study we apply them to a second-order CFA model.

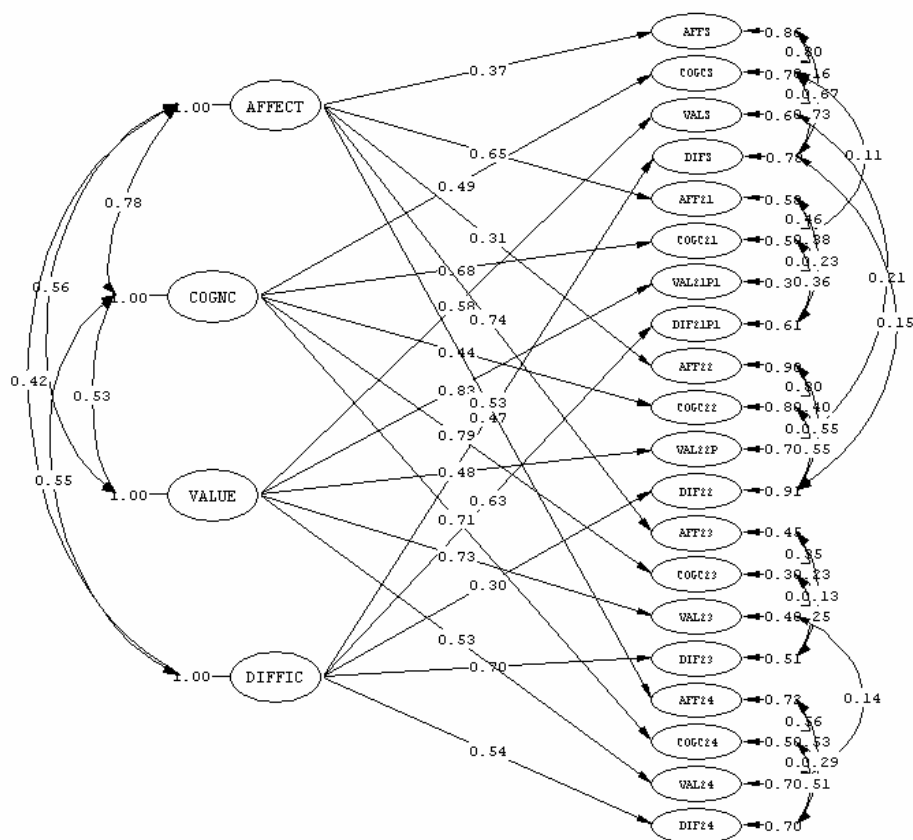


Figure 5-1. Second-order confirmatory factor model of achievement motivations. Values are standardized parameter estimates. All values shown are statistically significant, $p < .05$. AFFECT, Affect; COGNC, Cognitive Competence; VALUE, Value; DIFFIC, Difficulty

When contrasting Table 5.9 with Table 5.5, it is apparent that correlations of second-order motivations were rather similar to correlations of first-order motivations. One also notices that the correlations at the generic level reported in Table 5.9 were generally lower than at the subject-specific level

reported in Table 5.5. This indicates that at a generic level the different aspects of achievement motivation are more distinct than at the subject-specific level.

Table 5.10 shows how well second-order factors, representing the generic components of motivations, explain first-order factors. The last column of Table 5.10 contains the averages of explained variation per subject-matter area. On average, 35% of the variation of subject-specific achievement motivations is explained by generic motivations of the students. The degree of explanation of the first-order factors by generic components differs strongly by academic subject. This difference mirrors the Biglan classification. Only 24% and 15% of the variation in motivations for the hard subjects Statistics and Finance & Accounting, respectively, is explained by a generic motivational component, while the major part is subject-specific. In contrast, in the soft subjects roughly half of all variation is explained by the generic component.

Table 5-10. Explained % variation of first-order factors by second-order factors

	Affect	Cognitive Competence	Value	Difficulty	Subject average
Statistics	14	24	34	22	24
Business Strategy	41	46	67	39	48
Finance & Accounting	9	19	23	9	15
Marketing Management	56	64	53	49	56
Organization & HRM	28	50	31	30	35
Factor average	30	41	42	30	35

The last row of Table 5.10 shows, for each motivation, the average of the generic shares in variation of the first-order factors. The motivations Cognitive Competence and Value have the strongest generic components. This pattern holds across subject-matter areas. Moreover, for all academic specializations, the motivations Affect and Difficulty have the strongest subject-specific components. The results in Table 5.10 reinforce our conclusion from the previous CT, first-order CFA model. Disentangling Affect from a broader task value construct allows the motivation Value to incorporate most of the generic components, whereas Affect is mainly subject-specific.

3.5 Subject-specific structural equation models for achievement motivations and personality

From the results of the last section it is clear that the achievement motivations are not purely generic. Indeed, the results in Table 5.10 show that in most cases the subject-specific component dominates the generic one. This subject specificity leads us to the fourth research question. Specifically, we examined whether there are underlying factors that are not subject related but can still explain motivations for different academic specializations. This requires specification and test of a structural equation model that depicts the relation between the explanatory Big 5 personality factors and subject-specific achievement motivations. Further, it is important to estimate a structural equation model for each of the subjects separately: given that causal constructs are generic in nature, explanation of subject-specific achievement motivations can only take place through structural models being dependent on the subject. Table 5.1 indicates the five academic subjects, with the relevant sample sizes in the last column.

Table 5.11 contains fit indices of the structural models for personality and achievement motivations. The fit indices lead us to conclude that the structural models for each of the subjects fit the data reasonably well.

Table 5-11. Fit indices of structural models of personality and achievement motivations for five academic subjects

	χ^2	df	RMSEA	SRMR	GFI	NNFI	CFI	RFI
Statistics	991.41	497	.044	.053	.90	.94	.95	.89
Business Strategy	910.23	507	.048	.073	.86	.94	.95	.88
Finance &Accounting	867.54	505	.047	.076	.86	.95	.95	.88
Marketing Management	844.05	506	.048	.076	.84	.94	.95	.87
Organization & HRM	880.79	506	.052	.076	.83	.93	.93	.85

Table 5.12 presents the standardized estimates of structural path coefficients. Graphically, this is illustrated in Figure 5.2 for the subject Statistics. In particular, for the hard subjects Statistics and Finance & Accounting, personality factors Emotional Stability and Intellect explain the variation in motivations very well. This is in line with empirical studies relating goal setting to personality factors (Day, Radosevich, & Chasteen, 2003; Judge & Ilies, 2002; and Vermetten, Lodewijks, & Vermunt, 2001). In contrast, studies that consider the direct effect of personality on learning outcomes rather than on motivations, typically find that Conscientiousness

is the most important personality factor (De Raad & Schouwenburg, 1996; Matthews & Zeidner, 2005).

Table 5-12. Standardized estimates of structural path coefficients of structural model of personality and attitudes toward five academic subjects

	Affect	Cognitive Competence	Value	Difficulty
Statistics				
Extraversion				-0.14*
Agreeableness				
Conscientiousness				-0.19**
Emotional Stability	0.21***	0.14*		0.16*
Intellect	0.24***	0.44***	0.26***	0.30***
Business Strategy				
Extraversion	0.31***			
Agreeableness			0.18**	
Conscientiousness				
Emotional Stability				
Intellect		0.30***	0.16**	
Finance & Accounting				
Extraversion				
Agreeableness				
Conscientiousness				
Emotional Stability	0.34***	0.25***		0.18*
Intellect	0.18**	0.28***	0.30***	0.18**
Marketing Management				
Extraversion				
Agreeableness			0.26***	
Conscientiousness				
Emotional Stability	0.25***	0.19***		
Intellect	0.15*	0.23***		
Organization & HRM				
Extraversion	0.18**			
Agreeableness			0.24***	
Conscientiousness				
Emotional Stability	0.20***	0.15*		
Intellect		0.19**		

Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

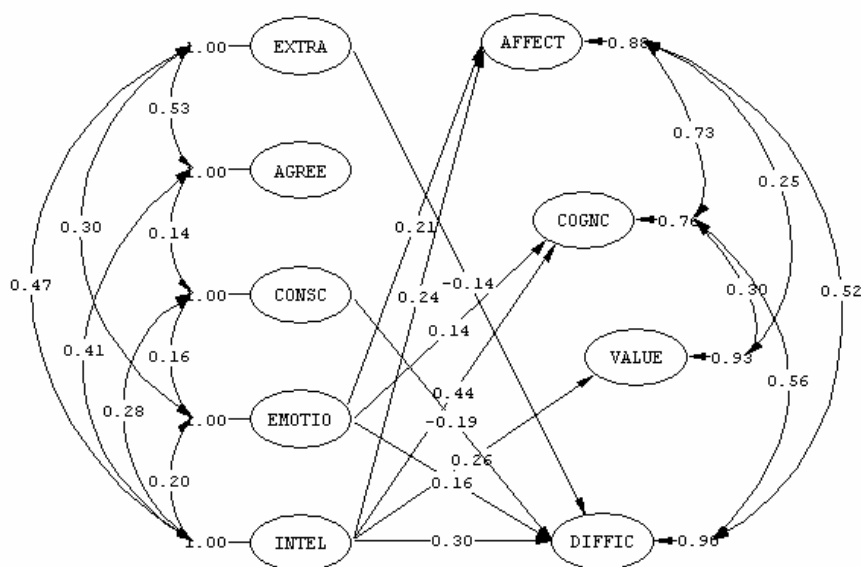


Figure 5-2. Structural equation model of personality and achievement motivations for statistics. Values are standardized parameter estimates. All values shown are statistically significant, $p < .05$. EXTRA, Extraversion; AGREE, Agreeableness; CONSC, Conscientiousness; EMOTIO, Emotional Stability; INTEL, Openness; DIFFIC, Difficulty; AFFECT, Affect; COGNC, Cognitive Competence; VALUE, Value

Marsh et al. (2006), for example, found Conscientiousness as the Big 5 factor most strongly related to mathematics self-concept. This is remarkable given that self-concept has strong affinity with the variable Cognitive Competence used in this study. Age differences between students in the Marsh et al. study (i.e., high school) and this and most other studies (i.e., university) might potentially explain why Intellect and, to a lesser extent, Emotional Stability, replace Conscientiousness as main predictors of perceived competence.

Comparing the results for the different content areas in Table 5.12, it is noticeable that for the hard subjects the achievement motivations in general, and specifically Difficulty and Affect, seem to be better explained than for the soft subjects. Earlier, we reported that Difficulty and Affect are the two motivations with the weakest generic component. Therefore, the conclusion is warranted that of all subject-specific achievement motivations, Affect and Difficulty for the hard subjects have predominantly a trait-like character.

4. CONCLUSION

The concept of latent variables and the corresponding methodology of structural equation modeling have in this study proven to be powerful tools in the research of subject-specific achievement motivations. It has allowed us to decompose motivations into generic and subject-specific components, and to investigate the relationship between achievement motivations and personality factors. In order to distinguish different components in subject-specific achievement motivations, this study adopted an affect-extended version of the expectancy-value model (Dauphinee et al., 1997; Hilton et al., 2004; Schau et al., 1995). Our data corroborated this affect extension, both in the domain of statistics and for four subjects in the business domain. In fact, given the purpose of this study to investigate generic versus subject-specific components in achievement motivations, the disentangling of affective and value related aspects from the broad task value construct proved to be very powerful: the affective and value-related constructs take opposite positions along the spectrum of generality versus subject-specificity.

Other empirical studies of achievement motivations over subject domains have mostly focused on children and adolescents in middle and high school, and their motivations for a range of broad domains such as math, reading, sports, or music (Wigfield & Eccles, 2000, 2002). Our study focuses on university students, and by consequence subject domains that are more congruent than in the earlier stages of schooling. In addition to investigating domain specificity at the level of subjects, it is also investigated in this article at the aggregated level of categories of subjects: hard versus soft subjects. The hard/soft decomposition generates patterns of achievement motivations that are very different, both with regard to the internal factor structure of motivations, as well as with regard to relationships with personality factors. But also at the academic subject level, even for subjects viewed as congruent, clear subject-specificity of achievement motivations is demonstrated.

Findings of our study are relevant for everyday educational practice. In the classroom, a student is quite often regarded as either motivated or not, irrespective of subject domain. Moreover, motivation is often regarded as an amorphous characteristic. It may contain different aspects, but these aspects are thought to be practically as well as empirically indistinguishable (Marsh & Yeung, 1996). Our research shows that achievement motivations, although containing an important generic component, are dominantly course-specific. Furthermore, achievement motivations are made up of

different aspects, some being rather stable, and others demonstrating stronger variability over subject-matter areas. Teaching and counseling should take the subject-specificity into account, in order to attain the maximum degree of efficiency. Students with motivational problems should be helped by the teacher teaching the specific subject where the problem is most imminent, in stead of counseling sessions detached from the teaching of academic content.

Our research also shows that major differences in the motivational structure between hard and soft subjects exist. Hard subjects stand out in low scores both on affective aspects and perceived lack of difficulty. At the same time, affect for these subjects is extremely strongly related to students' Cognitive Competence. Thus, to influence the affect for hard subjects one needs to consider all aspects of motivation. But, from this work it is apparent that of all subject-specific achievement motivations, Affect and (lack of perceived) Difficulty for the hard subjects have predominantly a trait-like character. It has also been shown that the score of Value for hard subjects is not strongly different from that of soft subjects, and that Affect and Value are relatively weakly correlated for hard subjects. Students seem to be able to combine the view that hard subjects, although not enjoyable and difficult, are still valuable. This implies that trying to convince students that hard subjects are valuable is not effective. Campaigns to influence students' choices in the direction of hard subject domains therefore do not seem promising. Also, because value has no distinctive influence, improving the compensation of careers in hard domains does not seem a fruitful alternative. Our research shows that a more promising approach to promote students' choice for hard subjects would be to work on students' cognitive competence.

4.1 Limitations, conclusion, and suggestions for future research

The power of the tool of latent variable modeling by structural equation methods has, however, its price. Specifically, these methods are rather data-intensive. In spite of the fact that we dispose of a large data set, the empirical limits are reached by the estimation of a second-order CFA model of both subject-specific and generic components of achievement motivations within the single-sample framework. Traditional measures of model complexity, expressed as cases per factor or parameter (Marsh et al., 1998), require a halt after the estimation of the second-order CFA. More

recently developed construct reliability measures (Gagné & Hancock, 2006) might suggest factor quality is adequate, but the limitation is here in the availability of appropriate benchmark values. The integration of the subject-specific full structural models developed in the last results subsection, into the second-order CFA of all subject achievement motivations, is certainly beyond the limits of our data set. Such an integrated second-order CFA, full structural model would enable the decomposition of all relations between achievement motivations and personality factors into a direct effect on subject-specific achievement motivations, and an indirect effect through the generic component of motivations.

This study has not addressed the relationship between achievement motivations and personality factors on the one hand, and course performance on the other. Such a study is a promising extension to this one, focusing on the question if the same personality factors that play an important role in explaining achievement motivations are also crucial in explaining course performance directly. Models of academic development such as the Model of Domain Learning of Alexander (1997) supports the importance of this line of inquiry as the study of cognitive and affective factors are examined simultaneously in relation to performance in a domain. Empirical research by Ackerman and colleagues (Ackerman, 2003a, 2003b; Ackerman, Bowen, Beier, & Kanfer, 2001) on the identification of trait complexes provides indirect evidence for the existence of domain specific relationships between stable achievement motivations, personality factors and learning performances. In a forthcoming study (Tempelaar, Schim van der Loeff, & Gijsselaers, 2007), we will address this issue.

REFERENCES

- Ackerman, P. L. (2003a). Aptitude complexes and trait complexes. *Educational Psychologist, 38*, 85-93.
- Ackerman, P. L. (2003b). Cognitive ability and non-ability trait determinants of expertise. *Educational Researcher, 32*, 15-20.
- Ackerman, P. L., Bowen, K. R., Beier, M. B., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology, 93*, 797-825.
- Alexander, P. A. (1992). Domain knowledge: Evolving themes and emerging concerns. *Educational Psychologist, 27*, 33-51.
- Alexander, P. A. (1997). Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In M. L. Maehr & P.

- R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 213-250). London, JAI Press.
- Bentler, P. M., & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16, 78-117.
- Biglan, A. (1973a). Relationships between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology*, 57, 204-213.
- Biglan, A. (1973b). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology*, 57, 195-203.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Burke, L. A., & Moore, J. E. (2003). A perennial dilemma in OB education: Engaging the traditional student. *Academy of Management Learning & Education*, 2, 37-52.
- Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.
- Coffman, D. L., & MacCallum, R. C. (2005). Using parcels to convert path analysis models into latent variable models. *Multivariate Behavioral Research*, 40, 235-259.
- Council of the European Union (2004). *Education & training 2010: The success of the Lisbon strategy hinges on urgent reforms*. Retrieved January 2, 2006, from <http://europa.eu.int/scadplus/leg/en/cha/c11071.htm>.
- Dai, D. Y., & Sternberg, R. J. (2004). Beyond cognitivism: Toward an integrated understanding of intellectual functioning and development. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual development and functioning* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum.
- Dauphinee, T. L., Schau, C., & Stevens, J. J. (1997). Survey of Attitudes Toward Statistics: Factor structure and factorial invariance for women and men. *Structural Equation Modeling*, 4, 129-141.
- Day, E. A., Radosevich, D. J., & Chasteen, C. S. (2003). Construct- and criterion-related validity of four commonly used goal orientation instruments. *Contemporary Educational Psychology*, 28, 434-464.
- De Raad, B., & Schouwenburg, H. (1996). Personality in learning and education: a review. *European Journal of Personality*, 10, 303-336.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual Review of Psychology*, 41, 417-440.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109-132.
- Gagné, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution property in confirmatory factor models. *Multivariate Behavioral Research*, 41, 65-83.

- Goldberg, L. R. (2005). *International personality item pool*. Retrieved January 2, 2006, from <http://ipip.ori.org/ipip/>.
- Gow, A. J., Whiteman, M. C., Pattie, A., & Deary, I. J. (2005). Goldberg's Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, 39, 317-330.
- Harris, M.B., & Schau, C. (1999). Successful strategies for teaching statistics. In S.N. Davis, M. Crawford & J. Sebrechts (Eds.), *Coming into her own: Educational success in girls and women* (pp. 193-210). San Francisco: Jossey-Bass.
- Hau, K. T., & Marsh, H. W. (2004). The use of item parcels in structural equation modelling: Non-normal data and small sample sizes. *British Journal of Mathematical Statistical Psychology*, 57, 327-351.
- Hilton, S. C., Schau, C., & Olsen, J. A. (2004). Survey of Attitudes Toward Statistics: Factor structure invariance by gender and by administration time. *Structural Equation Modeling: A Multidisciplinary Journal*, 11, 92-109.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55.
- Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology*, 87, 797-807.
- Kline, R.B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.
- Kolb, A.Y. & Kolb, D.A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning & Education*, 4, 193-212.
- Kulikowich, J. M., & DeFranco, T. C. (2003). Philosophy's role in characterizing the nature of educational psychology and mathematics. *Educational Psychologist*, 38, 147-156.
- Lawless, K. A. & Kulikowich, J. M. (2006). Domain knowledge and individual interest: The effects of academic level and specialization in statistics and psychology. *Contemporary Educational Psychology*, 31, 30-43.
- Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33, 181-220.
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality*, 74, 403-456.
- Marsh, H. W. & Yeung, A. S. (1996). The distinctiveness of affects in specific school subjects: An application of confirmatory factor analysis with the National Educational Longitudinal Study of 1988. *American Educational Research Journal*, 33, 665-689.
- Matthews, G. & Zeidner, M. (2004). Traits, states, and the trilogy of mind: an adaptive perspective on intellectual functioning. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on*

- intellectual development and functioning* (pp. 143-174). Mahwah, NJ: Lawrence Erlbaum.
- Pajares, F., & Schunk, D. (2005). The self and academic motivation: Theory and research after the cognitive revolution. In J. M. Royer (Ed.), *The cognitive revolution in educational psychology* (pp. 165-198). Greenwich, CT: Information Age Publishing.
- Paulsen, M. B., & Wells, C. T. (1998). Domain differences in the epistemological beliefs of college students. *Research in Higher Education, 39*, 365-384.
- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications*. Upper Saddle River, NJ: Pearson Education.
- Schau, C. (2003). *Students' attitudes: The "other" important outcome in statistics education*. Paper presented in the Joint Statistical Meetings, San Francisco, CA.
- Schau, C., Dauphinee, T. L., Del Vecchio, A., & Stevens, J. (1999). *Survey of attitudes toward statistics (SATS)*. Retrieved January 2, 2006, from <http://www.unm.edu/~cschau/downloadsats.pdf>.
- Schau, C.; Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement, 55*, 868-875.
- Schommer-Aikins, M., Duell, O. K., & Barker, S. (2003). Epistemological beliefs across domains using Biglan's classification of academic disciplines. *Research in Higher Education, 44*, 347-366.
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.
- Schunk, D. H. (2000). Coming to Terms with Motivational Constructs. *Contemporary Educational Psychology, 25*, 116-119.
- Snow, R. E., Corno, L., & Jackson, D. (1996). Individual differences in affective and conative domains. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 243-310). New York, NY: Macmillan Library reference.
- Sorge, C., & Schau, C. (2002, April). *Impact of engineering students' attitudes on achievement in statistics*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Tempelaar, D. T., Schim van der Loeff, S., & Gijssels, W. H., (2007). *Achievement motivations, personality factors and course performance in a range of academic subjects*. Maastricht: manuscript in preparation.
- Vermetten, Y. J., Lodewijks, H. G., & Vermunt, J. D. (2001). The role of personality traits and goals orientations in strategy use. *Contemporary Educational Psychology, 26*, 149-170.
- Weiner, B. (1992). *Human motivation: Metaphors, theories, and research*. Newbury Park, CA: SAGE Publications.
- Wigfield, A., Battle, A., Keller, L. B., & Eccles, J. S. (2002). Sex differences in motivation, self-concept, career aspirations and career choice: implications for cognitive development. In A. McGillicuddy-De Lisi & R. De Lisi (Eds.), *Biology, society, and behavior: The development of sex differences in cognition* (pp. 93-126). Westport: Ablex Publishing.

- Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, 12, 265-310.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy -- Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, 25, 68-81.
- Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Eds.), *Development of Achievement Motivation* (pp. 92-120). San Diego: Academic Press.

APPENDIX

In this Appendix the specification and the result, in terms of parameterisation, after the LISREL methodology of reduction is given of the various models that are discussed in Section 3. The purpose is, amongst others, to establish the degrees of freedom reported in the tables of that section.

For the models considered in Section 3 we use the following notation [c.f. Bollen(1989)]: For each subject domain j ($j = 1, \dots, 5$), briefly referred to as subjects, we use the measurement model

$$y^{(j)} = \Lambda_y^{(j)} \eta^{(j)} + \varepsilon^{(j)} \text{ with } Cov(\eta^{(j)}) = \Psi^{(j)} \text{ and } Cov(\varepsilon^{(j)}) = \Theta_\varepsilon^{(j)} \quad (A1)$$

where $y^{(j)}$ denotes the 9-dimensional vector containing the observed item-parcels and $\eta^{(j)}$ is the 4-dimensional vector of latent factors (motivations). Assuming a factor complexity of 1, i.e. assuming no cross-loadings, the matrix of factor loadings $\Lambda_y^{(j)}$ contains 9 unknown parameters. The covariance matrix of the errors of measurement $\varepsilon^{(j)}$, $\Theta_\varepsilon^{(j)}$, is a diagonal matrix and the covariance matrix of the latent variables $\Psi^{(j)}$ is unrestricted, where the diagonal elements are set equal to unity by means of normalization.

We also define the vector $y = (y^{(1)'} \dots y^{(5)'})'$ obtained by stacking the observations for all subjects in one vector (of dimension 45), and similarly $\varepsilon = (\varepsilon^{(1)'} \dots \varepsilon^{(5)'})'$. The stacked 20-dimensional vector of latent factors is denoted by $\eta' = (\eta^{(1)'} \dots \eta^{(5)'})'$.

In subsection 3.2 we use (A1), imposing the restriction that the covariance between *Difficulty* and *Value*, i.e. the (4,3)-element of $\Psi^{(j)}$, is a priori set equal to zero. This has the consequence that 23 parameters have to be estimated from the 45 distinct elements of the observed covariance matrix of the item-parcels, resulting in 22 degrees of freedom.

For the CT first^t-order factor model of subsection 3.3 the measurement model is

$$y = \Lambda_y \eta + \varepsilon \text{ with } Cov(\eta) = \Psi_{CT} \text{ and } Cov(\varepsilon) = \Theta_\varepsilon$$

The covariance matrix of the latent factors, Ψ_{CT} , has the structure

$$\Psi_{CT} = \begin{pmatrix} \Psi^{(1)} & \Psi^{(1,2)} & \dots & \Psi^{(1,5)} \\ \Psi^{(2,1)} & \Psi^{(2)} & \dots & \Psi^{(2,5)} \\ \vdots & \vdots & \ddots & \vdots \\ \Psi^{(5,1)} & \Psi^{(5,2)} & \dots & \Psi^{(5)} \end{pmatrix},$$

where the diagonal blocks $\Psi^{(j)}$ ($j = 1, \dots, 5$) are specified as in subsection 3.2, i.e. only imposing the restriction - in addition to normalization - that the covariance between *Difficulty* and *Value* is equal to zero. The off-diagonal blocks $\Psi^{(i,j)} = \Psi^{(j,i)}$ ($i, j = 1, \dots, 5; i \neq j$) are diagonal matrices. This yields 65 estimated parameters in the matrix Ψ_{CT} , four in each of the ten off-diagonal blocks, and five in each of the five diagonal blocks. Adding the two times 45 estimated parameters of Λ_y and Θ_ε , results in a total of 155 estimated parameters. Using the full covariance matrix of y , containing 1035 distinct elements leads to 880 degrees of freedom for the CT 1st-order factor model.

For the 2nd-order model for achievement motivation in subsection 3.4 the measurement model

$$y = \Lambda_y \eta + \varepsilon \text{ with } Cov(\varepsilon) = \Theta_\varepsilon$$

is complemented by a structural model, relating the latent (subject-specific) factors η to an underlying 4-dimensional vector ξ of exogenous latent factors (the generic motivations):

$$\eta = \Gamma \xi + \varsigma \text{ with } Cov(\xi) = \Phi \text{ and } Cov(\varsigma) = \Psi_{2F}$$

The standard notation for structural equation models is used here, necessitating the use of the same symbol for the covariance matrix of ζ as was used for η in the first-order models considered above, without risk of confusion (the equivalence is obtained by setting Γ equal to the zero matrix). The normalization is imposed on the matrix Φ , setting its diagonal elements equal to unity, and the (4,3)-element of Φ , giving the covariance between *Difficulty* and *Value*, is equal to zero. The initial structure of Ψ_{2F} is similar to that of Ψ_{cu} above (without normalization) allowing non-zero covariances between corresponding motivations for different subjects but, the resulting matrix has, with the exception of four elements [the (2,2)-element in the (1,2)- and (4,5)-blocks and the (3,3)- and (4,4)-element in the (1,3)-block] a block diagonal structure with, again, the covariances between *Difficulty* and *Value* in each diagonal block equal to zero, resulting in 49 estimated parameters in Ψ_{2F} . Together with the 2 times 45 estimated elements of Λ_y and Θ_ε , and the 5 estimated parameters in Φ , yields 144 parameters, estimated using the full covariance matrix of y . This leads to 891 degrees of freedom.

In subsection 3.5, Simultaneous Equation Models (SEM's) for each subject are considered: For $j = 1, \dots, 5$

$$y^{(j)} = \Lambda_y^{(j)} \eta^{(j)} + \varepsilon^{(j)} \text{ with } Cov(\varepsilon^{(j)}) = \Theta_\varepsilon^{(j)}$$

$$x^{(j)} = \Lambda_x^{(j)} \xi^{(j)} + \delta^{(j)} \text{ with } Cov(\xi^{(j)}) = \Phi^{(j)} \text{ and } Cov(\delta^{(j)}) = \Theta_\delta^{(j)}$$

$$\eta^{(j)} = \Gamma^{(j)} \zeta^{(j)} + \varsigma^{(j)} \text{ with } Cov(\zeta^{(j)}) = \Psi_{SM}^{(j)}.$$

here $x^{(j)}$ denotes the 25-dimensional vector of item-parcels resulting from the the Big-5 instrument and $\xi^{(j)}$ denotes the 5-dimensional vector of exogenous latent factors (personality). Also with regard to the model for the exogenous latent factors, a factor complexity of 1 is assumed and, hence, the matrix of factor loadings $\Lambda_x^{(j)}$ contains 25 unknown parameters. $\Psi_{SM}^{(j)}$ contains 5 estimated parameters (again the covariance between *Difficulty* and *Value* is equal to zero and the diagonal elements are normalized). $\Phi^{(j)}$ contains 8 unknown parameters (normalization constraints have been imposed on the diagonal elements). For all subjects this leads to 56 parameters to be estimated in the matrices $\Lambda_y^{(j)}$, $\Theta_\varepsilon^{(j)}$, $\Lambda_x^{(j)}$, $\Psi_{SM}^{(j)}$ and $\Phi^{(j)}$.

The full covariance matrix of the vector of item-parcels $(y^{(1)'}x^{(1)'})'$ of dimension 34 are used for estimation. Table A.1 presents the numbers of estimated parameters in the remaining matrices for each subject, along with the resulting degrees of freedom.

Table A-1. Number of estimated parameters and degrees of freedom per subject

	$\Gamma^{(j)}$	$\Theta_{\delta}^{(j)}$	Total estimated parameters	Degrees of freedom
Business Statistics	9	33	98	497
Business Strategy	4	28	88	507
Finance & Accounting	6	28	90	505
Marketing	5	28	89	506
Organization & HRM	5	28	89	506

The precise elements that are set equal to zero in the $\Gamma^{(j)}$ matrices can be reconstructed from Table 5.12.

Chapter 6

VARIATIONS IN STUDENT ACHIEVEMENT MOTIVATIONS*

1. INTRODUCTION

The research reported in this chapter extends the investigations into achievement motivations as a multi-dimensional (that is, differing over subjects and disciplines) and multi-faceted (that is, composed of different components) construct commenced in Chapter 5. By means of a second-order factor analytic model, we demonstrated in that chapter that subject-specific motivations contain a higher order, generic part and a part that is truly specific for the particular business subject. Through the estimation of subject-specific structural equation models explaining achievement motivations out of personality traits, we were able not only to provide evidence of multi-dimensionality, but also to demonstrate that a substantial part of motivations can be explained by stable constructs like personality traits. That last conclusion has important instructional implications: Teaching or, more generally, pedagogy is, in addition to its role of transferring knowledge per se, considered to have a role in forming motivations. To the extent that (parts of) achievement motivations are determined by non-malleable factors such as personality, these motivations are non-malleable themselves. The subject-specific structural equation models of Chapter 5 are the starting point of the analysis described in this

* This chapter is based on an abridged version of: Tempelaar, D. T., Schim van der Loeff, S., and Gijsselaers, W. H., Variations in achievement motivations over business subjects. Manuscript submitted to *Academy of Management Learning & Education*.

chapter. We extend the models in two directions: firstly, by adding gender as an additional background factor explaining achievement motivations and personality traits, and secondly, by adding course performance as an indicator of the outcomes of the learning process. The inclusion of gender in the model is again relevant with respect to the issue of malleability of motivations: gender stands for the most stable pole of background characteristics students bring into the educational process. But its marked stability is not the only reason to include gender in our model: learning processes are well documented to be gendered phenomena (see e.g. Alexander, 2006, and Byrnes, 2001, in a general context, or Rode et al., 2005, and Arbaugh, 2000, for examples in the context of management learning). The role of gender refers both the learning process itself, and determinants of that learning process. For example, Costa, Terracciano, and McCrae (2001), focusing on personality traits, report gender differences in these factors to be pervasive but moderate in size. The most consistent patterns they found are in Agreeableness, favoring women, and Emotional Stability, favoring men. Gender patterns found in achievement motivations are as consistent as those in personality factors, but somewhat more complex in nature. Firstly, both research in expectancy-value based motivations and research in students' self-concepts show differences in motivations according to gender, depending on the type of subject (Alexander, 2006; Marsh & Yeung, 1996; Shotic & Stephens, 2006; Wigfield & Eccles, 2000). These differences correspond with the position of subjects along the spectrum of the pure and hard versus applied and soft dimension of subjects: male pupils are in general more strongly motivated for mathematics and sciences, female pupils for languages. A second difference refers to the specific motivational facet: male pupils tend to be relatively more self-confident than female pupils, even for subjects where females tend to outperform males (Alexander; Marsh & Yeung; Wigfield & Eccles). A similar gender gap in self-perceptions is found by Shotic and Stephens (2006). In their recent study, they compare female and male students' self-efficacy ratings for 21 different computer tasks. Mean differences range from neutral in six tasks, to a significant gender gap favoring males in the other tasks.

The second extension of this study is the incorporation of course performance as an indicator of learning outcomes. It serves a clear purpose; how advanced the model describing the educational process might be, in the end one is primarily interested how well it explains the outcomes of that learning process. Integrating all components, the structure

of the model applied in this study is depicted in Figure 6.1. The Figure expresses that subjects are assumed to impact both the levels of motivational constructs, as the relationships between motivations and other variables.

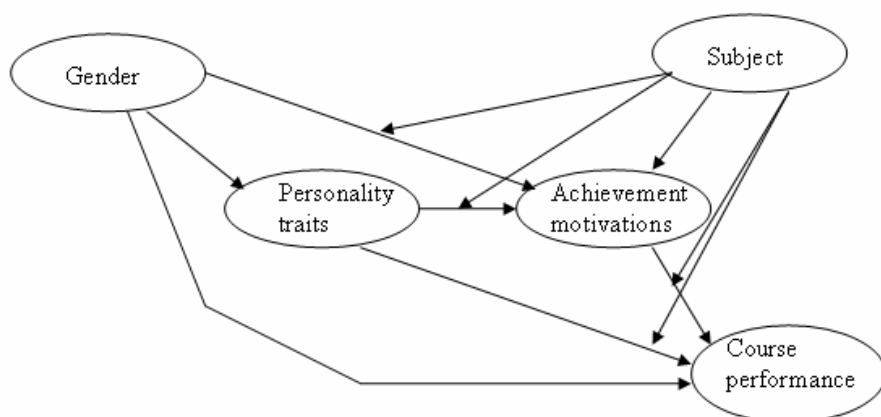


Figure 6-1. Model structure in this study

The important role that subjects and disciplines play in learning processes has been brought forward by other research. In the domain of business education, a prominent position is taken by studies analysing the relationships between business subjects and students' learning orientations as operationalised by Kolb's learning styles: see e.g. Kolb and Kolb (2005), Kolb (1984), and Loo's (2002) review article. The purport of this category of studies is that business disciplines have their own academic traditions and therefore pose varying demands on students. As a consequence, students of different business majors demonstrate different learning styles preferences. These studies are classroom studies, and typically compare different students in different majors. Our study too makes use of data collected in classroom, but uses a different setup than the one in most learning orientations studies. In order to disentangle true discipline related aspects in motivations from aspects related to students attracted by these disciplines, we investigate classroom practices of the same group of students in different intermediate courses. This setup was also chosen in a paper by Nijhuis, Segers, & Gijselaers (2007) in which discipline specificity of learning orientations is investigated. That study demonstrates that students use different learning strategies across three different business

courses. Differences in the variability of deep and surface learning were partially related to differences in perceptions of the discipline specific learning environments. In this study, we will investigate the role of motivational factors on learning business subjects within an otherwise similar design.

2. METHOD

From a methodological point of view, this study extends the study reported in Chapter 5. Therefore, the method section of this chapter is a shortened version of the method section of last chapter, with the exceptions of the instruments measuring course performance, and the methodology of comparing different groups. For a detailed description of all other issues, we refer to the method description included in Chapter 5.

2.1 Subjects, instruments, settings

Main participants in this longitudinal study are students in the program “International Business” of Maastricht University, who started their study in the academic year ‘99/’00. In their second year, those students take eight required intermediate level courses in core business and auxiliary subjects. In four of these second year courses, a large majority of students participated in a survey on subject achievement motivations. These were the three business courses Business Strategy, Marketing Management, and Organization & HRM, and the economics oriented Finance & Accounting course. Before entering the second year, these students, together with students of a parallel program “International Economics”, took required courses in the same subjects at an introductory level, and some more service courses. Our research started in one of these service courses: the first year course Quantitative Methods (QM), a required introduction into mathematics and statistics for all business and economics students, with approximately 1000 students participating.

In all five courses, achievement motivations for the respective subjects are measured with the Survey of Attitudes Toward Statistics (SATS) developed by Schau and co-authors (Dauphinee et al., 1997; Schau et al., 1995), and adaptations of this instrument for business subjects. Within the QM course, students’ personality traits were assessed with the 50-item

version of the IPIP Big5 factor unipolar markers (Goldberg, 2005; Gow, Whiteman, Pattie, & Deary, 2005).

2.2 Measurement of course performance

Of all five courses in which students achievement motivations for the relevant academic subject were assessed, students' raw scores in the final exam are available as indicator for course performance. Final exams in the courses Business Strategy, Finance & Accounting, Marketing Management, and Organization & HRM are written exams with objective items of true/false type. For the Quantitative Methods course, covering both business statistics and mathematics, multiple course performance indicators are available. In agreement with current assessment literature, learning performances are measured with a portfolio containing several instruments, each of them focusing on different aspects of the mastery of mathematical and statistical knowledge. Besides the before mentioned student project, the assessment instruments are:

- Final exams of the multiple choice format. The exam covers both statistics and mathematics; both parts are graded separately.
- Quizzes of multiple choice and short answer format for statistics. The quizzes allow students to achieve a bonus score. The level of the items is more basic than in the final exam, the main purpose being to stimulate students to spread their learning efforts evenly in time. It is hypothesized that the quiz score is stronger effort-based than the exam score.
- Weekly homework assignments of open type for both mathematics and statistics. The discussion of these assignments and the (partial) student solutions constitute the main agenda of the weekly, small-group, tutorial sessions. To get the discussions started, students were credited with some bonus for doing preparatory work on these assignments outside the tutorial group. Even more than the bonus for quizzes, these scores are assumed to be very strongly effort-based. Teaching assistants are explicitly instructed to assess the efforts put in by the students in trying to solve the homework problems, instead of assessing the correctness of the solution handed in.

The Quantitative Methods course is spread over three different terms, each with its own assessment. This implies that for each assessment type, three observations are available.

2.3 Statistical analyses for comparing gender and incorporating single indicators

Full structural models are composed of measurement parts and structural parts. In agreement with the two-step model building approach widely advocated (see, e.g., Schumacker & Lomax, 2004), the measurement models are estimated prior to the estimation of the structural model. Measurement models, both for achievement motivations and personality traits, can be characterized as correlated traits (CT), first-order confirmatory factor analysis (CFA) models. Gender effects in structural equation models are investigated through a sequence of invariance tests, whereby models excluding gender effects are compared with models including gender effect on the basis of differences in fit (Byrne, 1998). Two main types of a breakdown of gender invariance can be distinguished: where levels of constructs demonstrate a gender effect, but the structural model itself is still invariant, and the case where gender effects are present in both levels and the structural model based on the covariance structure. Before investigating gender effects in the structural equation context, the presence of gender effects in levels will be analyzed with independent samples t-tests. From those tests, percentage differences, t-values, and d-values or Cohen effect-sizes will be reported. The advantage of d-values over t-values is that they are independent of sample size. D-values larger or equal to 0.8, 0.5, and 0.2 respectively correspond to differences being large, medium, and small in size (Cohen, 1988).

Latent variable modeling assumes the existence of several manifest variables as indicators for each latent construct. In the case of course performance, data is restricted in all but one course to a single indicator: the grade in the final exam. In these cases, the modeling approach described in Bagozzi and Yi (1994) was followed: the factor loading of the single indicator was set to unity, and the residual variance to zero. For the subject business statistics, being part of a quantitative modeling course, several course performance variables are available, allowing the estimation of a latent factor model. Where course performance is measured with a single indicator, no assessment of the internal reliability of the instrument can be made. That is different for the portfolio of assessment instruments available for the QM course. Internal reliabilities (Cronbach α 's) for the five portfolio components range from 0.67 to 0.99. The reliability of the complete set of all indicators equals 0.78, indicating that internal reliability does not restrict the estimation of a measurement model for course performance. The

ecological validity of grades in the final exam stems from the circumstance that these data are the base of all academic pass/fail decisions.

3. RESULTS: MODELS FOR ACHIEVEMENT MOTIVATIONS ACROSS DOMAIN SUBJECTS

3.1 Genderedness of achievement motivations for academic subjects and its determinants

The present study investigates the existence of gender effects in students' achievement motivations in a range of academic subjects that demonstrate stronger similarity than subjects investigated in most research: all our subjects are part of a business curriculum. In this section, we will focus on the analysis of gender differences in levels of achievement motivations, leaving gender effects in relationships of motivations to other variables to the next section. A level analysis of data available in our study indicates the generic research findings to become corroborated in our study into business subjects. Table 6.1 describes the numerical outcomes of the several tests on gender effects, Figure 6.2 graphs the scale means of female and male students of all subject-specific achievement motivations. In Figure 6.2, the order of academic subjects is chosen such that the two pure subjects, the economics based Finance & Accounting and the mathematics based Business Statistics, are grouped on the left hand side, whereas the three applied subjects, the business oriented courses Marketing Management, Business Strategy, and Organization & HRM, are grouped on the right hand side.

From both exhibits, two conclusions are apparent. In the two pure subjects 'Stats' and 'Fin & Acc', male students express more positive levels of achievement motivations than female students. Differences are largest and strongly statistically significant for Affect and Cognitive Competence, but still substantive for Value and (lack of) Difficulty. The pattern for the applied subjects 'MarkMan', 'BusStrat', and 'OrgHRM', is mixed: gender differences are much smaller, and in both directions, indicating that, relative to male students, female students have higher motivation levels for applied business subjects, than for pure non-business subjects.

Table 6-1. Scale means for female and male students, percentage gender difference (%Diff) with male score as basis, and t-value and d-value for gender differences of achievement motivations

	Female mean	Male mean	%Diff	<i>t</i> -value	<i>d</i> -value
Affect (1-7):					
Business Statistics	4.28	4.58	-6.50	-4.88	-0.32
Business Strategy	5.47	5.56	-1.60	-1.06	-0.12
Finance & Accounting	3.72	4.25	-12.41	-3.93	-0.48
Marketing Management	5.43	5.43	0.07	0.04	0.01
Organization & HRM	5.52	5.21	5.96	3.57	0.46
Cognitive Competence (1-7):					
Business Statistics	4.57	4.94	-7.40	-7.07	-0.47
Business Strategy	5.50	5.75	-4.36	-3.34	-0.38
Finance & Accounting	4.19	4.68	-10.46	-4.40	-0.53
Marketing Management	5.34	5.49	-2.67	-1.52	-0.20
Organization & HRM	5.33	5.28	0.94	0.57	0.07
Value (1-7):					
Business Statistics	5.03	5.07	-0.64	-0.68	-0.05
Business Strategy	5.62	5.53	1.70	1.21	0.14
Finance & Accounting	4.70	4.94	-4.84	-2.23	-0.27
Marketing Management	5.60	5.39	3.95	2.27	0.30
Organization & HRM	5.39	5.09	5.96	3.22	0.41
Difficulty (1-7):					
Business Statistics	3.43	3.61	-4.81	-4.05	-0.27
Business Strategy	4.32	4.39	-1.66	-0.90	-0.10
Finance & Accounting	2.75	2.88	-4.58	-1.61	-0.20
Marketing Management	4.19	4.28	-2.00	-1.01	-0.13
Organization & HRM	4.53	4.53	0.08	0.04	0.01

The next table, Table 6.2, contains similar descriptive statistics for personality traits and course performances.

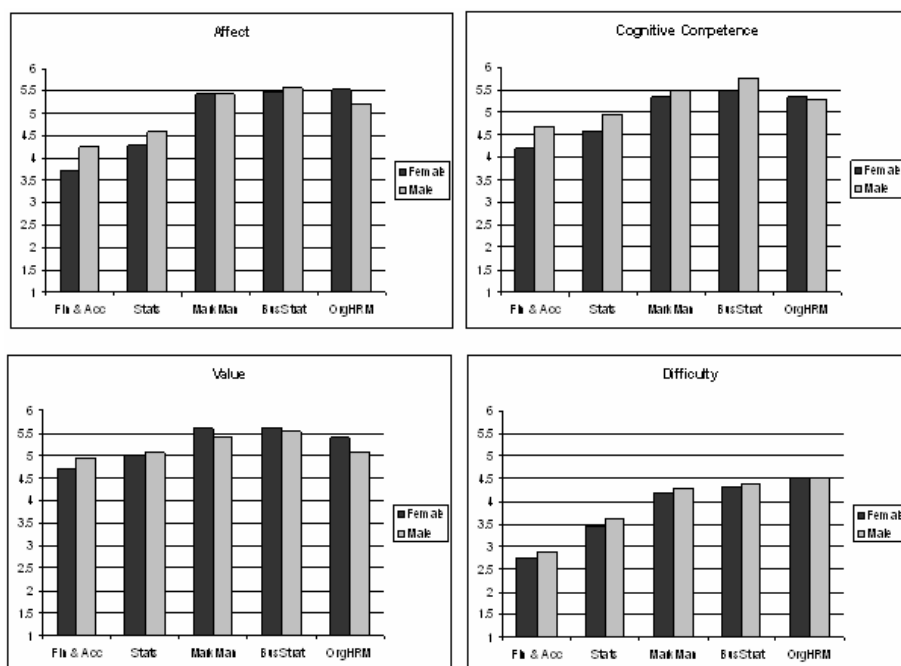


Figure 6-2. Mean attitudes scores of subjects Finance & Accounting, Statistics, Marketing Management, Business Strategy, and Organization & HRM, for female and male students

Combining the results of Table 6.1 and Table 6.2, some further interpretations are noticeable. The largest gender differences favoring male students are found in the self-efficacy scale Cognitive Competence. However, this self-reported optimism, relative to the other gender, does not translate itself into better academic performance: female students outperform male students in all course performance measures, except two partial course performances: the two final exam components of the QM course. The several partial course performance measures for QM demonstrate in themselves a clear pattern: the strongly effort-based indicators for Homework demonstrate a gender effect favoring female students, the Quiz indicator a less pronounced gender effect, and the cognitive based Exam indicators demonstrate a gender effect with opposite direction.

Table 6-2. Scale means for female and male students, percentage gender difference (%Diff) with male score as basis, and *t*-value and *d*-value for gender difference, of achievement motivations, personality traits, and course performance measures

	Female mean	Male mean	%Diff	<i>t</i> -value	<i>d</i> -value
Big5 Personality (1-5):					
Extraversion	3.33	3.47	-4.03	-2.70	-0.24
Agreeableness	3.86	3.67	5.18	5.28	0.47
Conscientiousness	3.42	3.38	1.18	0.94	0.08
Emotional Stability	3.06	3.35	-8.66	-5.68	-0.51
Intellect	3.29	3.50	-6.00	-5.30	-0.47
Course performance (1-10):					
QM: Math Final Exam	5.52	5.95	-7.13	-2.70	-0.24
QM: Statistics Final Exam	6.62	6.97	-5.00	-2.80	-0.25
QM: Statistics Quiz	5.75	5.63	2.22	1.05	0.09
QM: Math Homework	6.58	6.29	4.72	2.89	0.26
QM: Statistics Homework	5.11	4.80	6.52	3.03	0.27
Business Strategy	6.25	6.00	4.22	1.54	0.20
Finance & Accounting	5.81	5.73	1.50	0.36	0.05
Marketing Management	5.58	5.18	7.63	1.98	0.27
Organization & HRM	6.59	6.00	9.91	2.73	0.37

A close inspection of both tables suggest that where both gender differences in self efficacy and course performance fluctuate over subjects, the combined effect of the two is rather stable. Such a combined effect of relative self efficacy and relative course performance may be regarded as a numerical approximation of the extent to which male students are ‘over-confident’ about their academic capabilities, or equivalently, female students underestimate their true capabilities. The most simple version of such a criterion of over-confidence is the difference between percentage gender difference in Cognitive Competence and percentage gender difference in course performance. That measure is indeed remarkably stable over subjects: Business Statistics 9.62%, Business Strategy 8.58%, Finance & Accounting 11.96%, Marketing Management 10.30%, and Organization & HRM 8.97% (for Business Statistics, the quiz indicator was chosen to measure performance, the indicator taking the median position in self-efficacy). So whilst self-efficacy and performance are both subject dependent and gendered, the combined effect expressed as over-confidence is not subject dependent, but only gendered, and demonstrates minor fluctuations around a substantive 10% over confidence of males

relative to females. Finally, self-reported personality traits levels for males exceed those for females for Extraversion, Emotional Stability, and Intellect. The exception to the pattern is Agreeableness, with scores for Conscientiousness being rather similar.

All in all, gender effects in subject-specific achievement motivations appear to be substantial. Gendered student characteristics constitute a clear exponent of highly stable determinants of the learning process. The next section introduces a second set of student characteristics into the analysis with a dominant stable nature: personality traits.

3.2 Structural models for personality traits, achievement motivations, and course performance in a range of subjects

Finding different levels of achievement motivations for a range of business subjects (and for personality factors, and course performances) in female and male students does not imply that measurement models for these constructs or structural models cannot be the same for female and male students. In this section, after having established gender non-invariance in levels, we will focus on the analysis of gender invariance in terms of covariance structures. In agreement with the two-step approach of estimating structural equation models, measurement models for personality factors, subject-specific achievement motivations, and course performance were estimated prior to the estimation of the full structural model. We will focus on reporting of the full models in this contribution, and refer to Chapter 5 for characteristics of the measurement models. All measurement models demonstrate adequate fit, and the measurement models for achievement motivations are non-invariant over subjects: Different subjects are described by different factor models, whereby especially the correlations between motivations depend on subjects. For that reason, full structural models were estimated for all subjects separately. Table 6.3 contains fit indices of all five subject models, indicating that all subject models fit rather well, given the size of the models (χ^2/df ratio's are all below two).

Table 6-3. Fit indices of four-factor confirmatory factor models of achievement motivation for five academic subjects

	χ^2	df	RMSEA	SRMR	GFI	NNFI	CFI	RFI
Business Statistics	1997	1043	.044	.062	.86	.95	.95	.89
Business Strategy	35	22	.041	.036	.98	.99	.99	.97
Finance & Accounting	892	540	.050	.080	.83	.94	.94	.85
Marketing Management	849	539	.047	.077	.83	.94	.95	.85
Organization & HRM	912	538	.052	.076	.82	.92	.93	.83

In the remainder of this section, all five subject models will be discussed, starting with the Business subjects. Figure 6.3 describes the structural equation model for the subject Organization & HRM. The left hand side of the exhibit contains the measurement model of students' personality traits; that part is identical in all subject models. Curved arrows in the left part indicate that latent traits are correlated, with especially Intelligence demonstrating strong correlations with Extraversion and Agreeableness. Not all pairs of traits are however correlated: correlations between Extraversion and Conscientiousness, and between Agreeableness and Emotional stability, are not statistically different from zero, and are therefore restricted to zero in all final models. The right hand side of Figure 6.3 contains the measurement model of students' achievement motivations, and the single indicator for course performance. Again, latent motivation factors are correlated, except for one pair: Value appears to be unrelated to (lack of perceived) Difficulty, for all subjects. Strong correlations are present amongst the three constructs Affect, Cognitive Competence, and Value, and in the pair Cognitive Competence and (lack of) Difficulty.

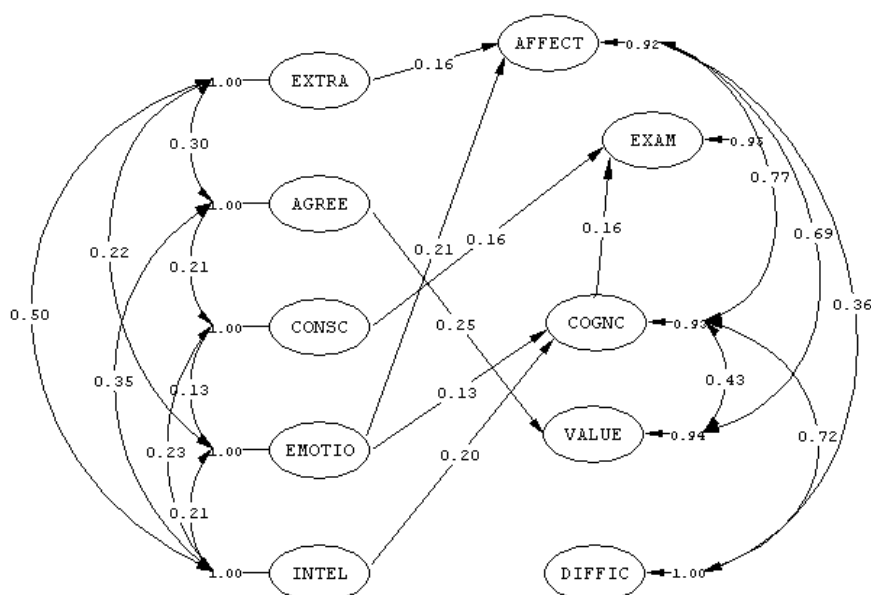


Figure 6-3. Structural equation model for Organization & HRM. Values are standardized parameter estimates. All values shown are statistically significant, $p < .05$. EXTRA, Extraversion; AGREE, Agreeableness; CONSC, Conscientiousness; EMOTIO, Emotional Stability; INTEL, Openness; DIFFIC, Difficulty; AFFECT, Affect; COGNC, Cognitive Competence; VALUE, Value; EXAM, course performance

The structural part of the model contains five statistically significant paths from personality to motivations. The strongest relationships are those between Intellect and the self-efficacy variable Cognitive Competence, between Agreeableness and task Value, and between Emotional Stability and Affect. If we compare the model for Organization & HRM with the second model, for Marketing Management depicted in Figure 6.4, there exists a strong degree of similarity. Measurement models are similar, and the same three paths are found as dominant relationships between personality factors and motivations. However, other paths are different; e.g. in Organization & HRM, a path from Extraversion is present, which is absent in the Marketing Management model. The strongest difference is however in the explanation of course performance: where both Conscientiousness and Cognitive Competence explain Exam for Organization & HRM, it is only Intellect that explains course performance for Marketing Management.

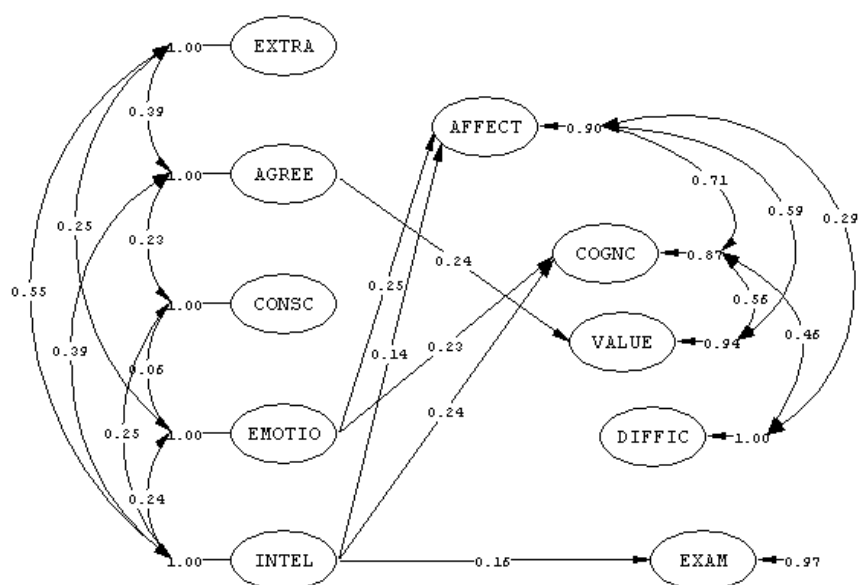


Figure 6-4. Structural equation model for Marketing Management. Values are standardized parameter estimates. All values shown are statistically significant, $p < .05$. EXTRA, Extraversion; AGREE, Agreeableness; CONSC, Conscientiousness; EMOTIO, Emotional Stability; INTEL, Openness; DIFFIC, Difficulty; AFFECT, Affect; COGNC, Cognitive Competence; VALUE, Value; EXAM, course performance

Extending the analysis to the third business subject, Business Strategy generates a similar pattern: see Figure 6.5. Measurement models and the three dominant paths between Intellect and the self-efficacy variable Cognitive Competence, between Agreeableness and task Value, and between Emotional Stability and Affect remain unchanged. However, the Business Strategy model contains weaker relationships in the sense that beyond this common part, only a fourth weak path is present between Intellect and Value, but no path providing any explanation for course performance indicator EXAM. For Business Strategy, there appears to be no statistically significant relationship between personality factors and motivations as explanatory variables of course performance. In all models, task difficulty, expressed as lack of perceived Difficulty, is the only motivational variable being unrelated to any personality trait.

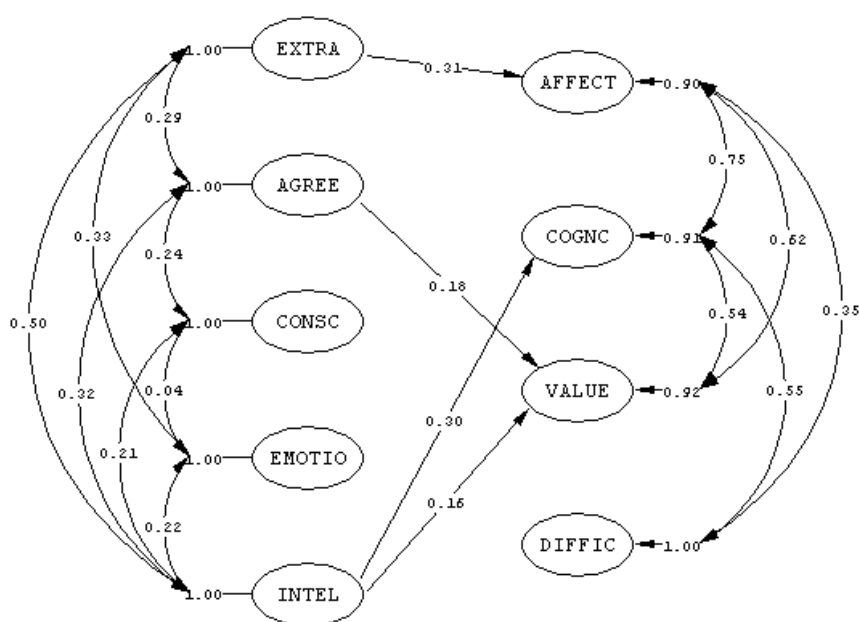


Figure 6-5. Structural equation model for Business Strategy. Values are standardized parameter estimates. All values shown are statistically significant, $p < .05$. EXTRA, Extraversion; AGREE, Agreeableness; CONSC, Conscientiousness; EMOTIO, Emotional Stability; INTEL, Openness; DIFFIC, Difficulty; AFFECT, Affect; COGNC, Cognitive Competence; VALUE, Value

When switching to the pure subjects, the models depicted in Figure 6.6 for Finance and Accounting, and in Figure 6.7 for Business Statistics, demonstrate clear departures from the models for the business subjects. A first difference is in the right hand side of the models, the measurement model of achievement motivations. The differences are concentrated in the correlations between latent factors Value and Affect, and Value and Cognitive Competence. In the last subsection, we have noticed that subject-differences in Affect and Cognitive Competence are much larger than those in Value, and all in the same direction: pure subjects are characterized with lower motivational levels than applied subjects. To this we can now add that correlations between Value and Affect are much lower for pure subjects (0.48, 0.24) than for applied subjects (0.69, 0.59, 0.62). And similarly, correlations between Value and Cognitive Competence are lower for pure subjects (0.48, 0.30) than for applied subjects (0.43, 0.56, 0.54), be it that this effect is less univocal. Apparently, the liking of a subject and valuing it, and feeling competent in it, are more or less two sides of one

medal in the case of the business subjects, whereas students are able to attach value to the pure subjects without liking them. The largest subject differences were found in (lack of) Difficulty; not surprisingly, this variable is not correlated at all with Value. So for all subjects, students can value the subject, both in case they regard the subject as difficult, or not.

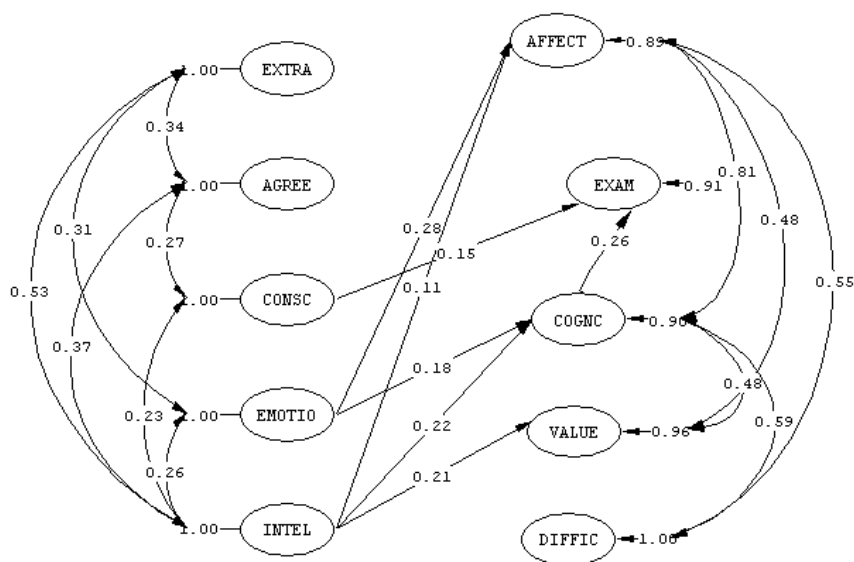


Figure 6-6. Structural equation model for Finance & Accounting. Values are standardized parameter estimates. All values shown are statistically significant, $p < .05$. EXTRA, Extraversion; AGREE, Agreeableness; CONSC, Conscientiousness; EMOTIO, Emotional Stability; INTEL, Openness; DIFFIC, Difficulty; AFFECT, Affect; COGNC, Cognitive Competence; VALUE, Value; EXAM, course performance

In terms of the relationship between personality traits and achievement motivations, the Intelligence constructs takes a far stronger role in the pure subjects, than in the applied subjects. Whereas the path to Cognitive Competence is the only robust one in the business subjects, Intelligence explains all motivations for Business Statistics, and all but Difficulty for Finance & Accounting. For Business Statistics, Intelligence is in fact the only predictor of motivations, except for path between Conscientiousness and (lack of) Difficulty, which path has a negative coefficient, indicating that the more conscientious students regard statistics as more demanding than the less conscientious students. Another difference between pure and

applied subjects is that Extraversion and Agreeableness are unrelated to any motivational construct for the pure subjects, different from the applied ones.

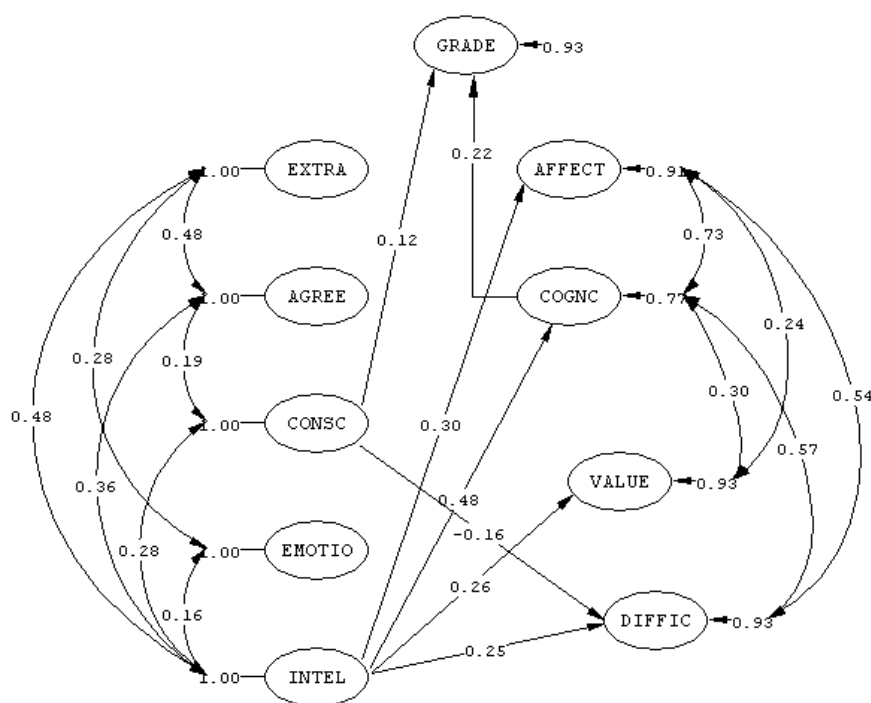


Figure 6-7. Structural equation model for Business Statistics. Values are standardized parameter estimates. All values shown are statistically significant, $p < .05$. EXTRA, Extraversion; AGREE, Agreeableness; CONSC, Conscientiousness; EMOTIO, Emotional Stability; INTEL, Openness; DIFFIC, Difficulty; AFFECT, Affect; COGNC, Cognitive Competence; VALUE, Value; GRADE, Overall grade

Figure 6.7 depicts the structural equation model for Business Statistics when using the overall Grade as a single course performance indicator. Data used to estimate this model is comparable to data used in the estimation of the other subject-specific models. Grade is a weighted average of the scores achieved in the several portfolio components, and is different from Exam, indicating the scores achieved in the final exam (be it that the weights of exam scores in Grade are much higher than those of homework and quizzes). The model depicted in Figure 6.8 makes use of the complete richness of the portfolio of course performance assessments

Figure 1 is a path diagram of the second-order confirmatory factor analysis model. The diagram shows five latent variables (EXTRA, AGREE, CONSC, EMOTIO, INTEL) and their relationships with observed variables (MATHHOME, STATHOME, STATQUIZ, MATHEXAM, STATEXAM, AFFECT, COGNC, VALUE, DIFFIC). Standardized path coefficients are provided for each relationship, along with correlations between latent variables.

Latent Variables and Observed Variables:

- EXTRA (Latent Variable)
- AGREE (Latent Variable)
- CONSC (Latent Variable)
- EMOTIO (Latent Variable)
- INTEL (Latent Variable)
- MATHHOME (Observed Variable)
- STATHOME (Observed Variable)
- STATQUIZ (Observed Variable)
- MATHEXAM (Observed Variable)
- STATEXAM (Observed Variable)
- AFFECT (Observed Variable)
- COGNC (Observed Variable)
- VALUE (Observed Variable)
- DIFFIC (Observed Variable)

Standardized Path Coefficients:

- EXTRA → MATHHOME: 0.88
- EXTRA → STATHOME: 0.91
- EXTRA → STATQUIZ: 0.93
- EXTRA → MATHEXAM: 0.90
- EXTRA → STATEXAM: 0.93
- AGREE → MATHHOME: 0.32
- AGREE → STATHOME: 0.32
- AGREE → STATQUIZ: 0.13
- AGREE → MATHEXAM: 0.22
- AGREE → STATEXAM: 0.12
- CONSC → MATHHOME: 0.32
- CONSC → STATHOME: 0.32
- CONSC → STATQUIZ: 0.13
- CONSC → MATHEXAM: 0.22
- CONSC → STATEXAM: 0.12
- EMOTIO → AFFECT: 0.91
- EMOTIO → COGNC: 0.74
- EMOTIO → VALUE: 0.74
- EMOTIO → DIFFIC: 0.74
- INTEL → AFFECT: 0.91
- INTEL → COGNC: 0.74
- INTEL → VALUE: 0.74
- INTEL → DIFFIC: 0.74

Correlations between Latent Variables:

- EXTRA ↔ AGREE: 0.48
- EXTRA ↔ CONSC: 0.29
- EXTRA ↔ EMOTIO: 0.38
- EXTRA ↔ INTEL: 0.28
- AGREE ↔ CONSC: 0.17
- AGREE ↔ EMOTIO: 0.16
- AGREE ↔ INTEL: 0.16
- CONSC ↔ EMOTIO: 0.16
- CONSC ↔ INTEL: 0.16
- EMOTIO ↔ INTEL: 0.16

Figure 6-8. Structural equation model for Business Statistics. Values are standardized parameter estimates. All values shown are statistically significant, $p < .05$. EXTRA, Extraversion; AGREE, Agreeableness; CONSC, Conscientiousness; EMOTIO, Emotional Stability; INTEL, Openness; DIFFIC, Difficulty; AFFECT, Affect; COGNC, Cognitive Competence; VALUE, Value; MATHHOME, Homework mathematics; STATHOME, Homework statistics; STATQUIZ, Quiz statistics; MATHEXAM, Final exam mathematics; STATEXAM, Final exam statistics

The most powerful predictor is the personality trait Conscientiousness. However, it is especially powerful for the strongly effort dependent homework scores, already less powerful for the quiz scores, and least

powerful for the score in the final exam (with math final exam score even being unrelated to it). Math homework is even determined in two independent ways by Conscientiousness: directly, and indirectly through Difficulty: Conscientious students regard statistics as more demanding, and students who regard the subject as more demanding, score higher on math homework. It is remarkable how different the two most important personality traits behave in all subjects models. Conscientiousness has, except for the weak relationship with Difficulty in the model for Statistics as explained above, no impact on any motivational variable, but at the same time is the strongest direct predictor of course performance in the several subjects. In contrast, Intellect is the personality traits with strongest relationships to motivations but has, except for the subject Marketing Management, no direct impact on course performance.

In terms of explained variation, the relationships between gender and personality factors and achievement motivations, and between all these three factors and course performance, is of moderate strength. Gender explains a median 4% of variation in subject-specific achievement motivations, with a maximum of 7% for Cognitive Competence in Finance & Accounting. On top of that, personality traits explain a further median 8% of variation in motivations, with a maximum of 22% for Cognitive Competence in Business Statistics. These data strengthen the conclusion of the first empirical section: due to the fact that a substantial part of achievement motivations is determined by stable factors as gender and personality, there seems to be only restricted potential for pedagogy. Differences between the several structural models estimated for different subjects, point to a further restriction of pedagogy: that pedagogy has to be course-specific, in order to accommodate course-specific relationships between personality factors, achievement motivations, and course performance.

3.3 Gender invariance of structural models for achievement motivations and course performance

After establishing strong gender effects in levels of achievement motivations and course performance, and finding that structural equation models that explain achievement motivations and course performance out of personality factors, the last research question refers to the possibility that these subject-specific structural models are gender-invariant. Investigations into gender-invariance of structural models starts with invariance checks applied to measurement models of Big 5 personality factors, and the

measurement models of course specific achievement motivations. Since the size of the models makes full reporting on all invariance tests unattractive (the structural equation model for Business Statistics e. g. contains 149 free parameters, see Appendix, requiring 149 invariance tests on pairs of parameter estimates), we will discuss the outcomes in a more qualitative manner. The structural equation models appear to be gender-invariant, except for two motivation models: Finance & Accounting, and Marketing Management, demonstrate a break-down of gender invariance in the correlation matrix of latent achievement motivations. In Finance & Accounting, the rejection of invariance is marginal, and concentrated in two correlations: for male students, the correlation between Value and Cognitive Competence and Value and Affect is much stronger than for females than for males. This lower correlation for males contains the clue to explain how the large gender gap in Value is compatible with smaller gender gaps in Affect and Cognitive Competence. Female students tend to value the subject especially if they feel cognitive competent and possess positive affect; in male students, these two circumstances are much less a condition for positive valuation. The breakdown in gender-invariance for Marketing Management is primarily in the correlation between Affect and Cognitive Competence. In females, that correlation is so strong as to make the two factors empirically non-distinguishable. In males, the correlation is strong (0.70), but clearly distinguishable.

Testing for gender-invariance in the complete structural model results in similar outcomes. Structural models for subjects Business Statistics, Business Strategy, and Organization & HRM are gender-invariant, except for obvious differences in latent means. The variation by gender of the motivations measurement models for the subjects Finance & Accounting, and Marketing Management, also has an impact on the structural models for these subjects. For Finance & Accounting, the breakdown of invariance is limited to the correlation matrix of latent motivation factors. Gender invariance cannot be rejected for path coefficients relating personality traits with achievement motivations for this subject. For the subject Marketing Management, however, gender invariance in both the motivations correlations, and the paths coefficients must be rejected. It appears that gender specific path coefficient matrices are complementary: all five paths coefficients of the model estimated on the complete sample, see Figure 6.3, are either significant in the female sub-sample, but not in the male sub-sample, or vice versa. For males, Emotional Stability is the single personality factor predicting Affect and Cognitive Competence. Both these

variables are predicted by Intellect for females, as is Value by Agreeableness.

Taking both types of gender effects altogether, the gender effect in levels clearly dominates the gender effect in terms of covariance structures. All subjects demonstrate strong level effects, whereas three out of five subjects are characterized by invariant structural models. Two out of five structural models exhibit significant gender effects; however, these are located in a few parameters within the covariance matrix of latent motivation constructs. Taking differences in levels and those in covariance structures together, our studies suggest that substantial gender effects are present, but that these primarily show up as differences in levels, not as differences in models structures.

4. DISCUSSION

Gender patterns in levels of achievement motivations, personality factors, and course performance, corroborate generic findings in learning research. There exists a strong gender effect in attitudes toward subjects that is in size depending upon the nature of the subject: gender differences whereby males report more positive attitudes than females are stronger for the pure subjects, than for the applied subjects. Most remarkable are gender differences in self-perceived competence, and males reporting high levels of attitudes while at the same time being outperformed by females. Gender effects are substantive, sometimes up to ten percentage points. Although gender effects in both motivations and course performance is subject-specific, the sum of percentage point gender difference in self-perceived competence, and percentage point underachievement, renders a very stable 10% for all academic subjects, suggesting that males' overconfidence is a stable characteristic independent of the academic subject. The circumstance that subject-specific competence beliefs may be the product of a constant 'overconfidence' gender gap, and subject-specific mastery levels, has important instructional implications. In their study into self-perceptions in a range of computer tasks, Shotic and Stephens (2006) conclude that gender gaps are a function of the specific task, and that studies focusing on comparing academic tasks and academic subjects should suggest where (business) schools encounter the biggest need to strengthen female students' confidence levels. But if our finding of a stable overconfidence level can be extended to the area of computer skills, it is not that obvious that the focus is so strongly on the selection of tasks and

subjects. Male students' over-confidence, or female students' relative underestimation of capabilities, might be a very generic phenomenon, rather unrelated to subject.

Gendered student characteristics constitute an outstanding example of non-malleable student-related factors that determine learning processes. Beyond gender, we have studied another determinant of learning that is that stable as to regard it as non-malleable: personality traits. Both factors together explain 10%-20% of variation in achievement motivations, with strong variation over type of motivation and academic subject. The set of stable student characteristics determining learning can easily be extended beyond gender and personality. Preferred learning styles, discussed in the introduction of this paper, is only one further example of a very stable student characteristic having a pronounced impact on student learning. As a consequence, student learning in business subjects is to a substantial extent depending on non-malleable student characteristics. This poses severe constraints with regard to the attainability of instructional aims related to improving student motivations.

The remaining part of the spectrum of student-related determinants of the learning process consists of malleable factors. Making use of the outcomes of Chapter 5, we can distinguish between characteristics that are subject-specific, and those that are generic for a range of subjects. To distinguish between these two components is of great relevance to recent discussions on 'integrative pedagogy' in business education. In an attempt to meet the critics of the 'break down the silos' movement (Stover et al. 1997; Bennis & O'Toole, 2005), implying that programs in (undergraduate) business schools focus too much on isolated business functions and lack cross-functional integration, Campbell, Heriot and Finney (2006) recently propose to distinguish between a focus on pedagogy and a focus on curriculum. They argue that the problem with business education is not with *what* is taught in business courses, but with *how* it is taught. Therefore, they conclude, the proper response to criticisms of business programs is not to change the curriculum, but rather to adopt a more integrative pedagogy. But such an integrative pedagogy is strongly built on the assumption that important determinants of students learning, such as learning motivations, can be influenced on a generic, not subject-specific, level. However, the decomposition of motivational factors into generic and subject-specific components makes also clear that the perspectives of an integrated pedagogy are limited: the far greater part of student motivations is of course specific type, and only a modest part is of generic type. In order to improve

students' motivations in the most malleable part of the spectrum, those that are strongly context dependent, it is hard to imagine that a pedagogy that does not address these context specific issues can play a large role. For that reason, breaking down the silos has indeed its restrictions, both from the curriculum, and the pedagogic perspective.

REFERENCES

- Alexander, P. A. (2006). *Psychology in learning and instruction*. Upper Saddle River, N.J.: Pearson.
- Arbaugh, J. B. (2000). An exploratory study of the effects of gender on student learning and class participation in an internet-based MBA course. *Management Learning*, 31, 503-519.
- Bagozzi, R. P., & Yi, Y. (1994). Advanced topics in structural equation models. In R. P. Bagozzi (Ed.), *Advanced methods in marketing research*. Cambridge: Basil Blackwell.
- Bennis, W.G., & O'Toole, J. (2005). How business schools lost their way. *Harvard Business Review*, May, 1-9.
- Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.
- Byrnes, J. P. (2001). *Cognitive development and learning in instructional contexts* (2nd ed.). Boston, MA: Allyn and Bacon.
- Campbell, N. D., Heriot, K. C., & Finney, R. Z. (2006). In defense of silos: An argument against the integrative undergraduate business curriculum. *Journal of Management Education*, 30, 316-332.
- Cohen, J. (1998). *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Costa, P. T. Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality & Social Psychology*, 81, 322-331.
- Dauphinee, T. L., Schau, C., & Stevens, J. J. (1997). Survey of Attitudes Toward Statistics: Factor structure and factorial invariance for women and men. *Structural Equation Modeling*, 4, 129-141.
- Goldberg, L. R. (2005). *International personality item pool*. Retrieved January 2, 2006, from <http://ipip.ori.org/ipip/>.
- Gow, A. J., Whiteman, M. C., Pattie, A., & Deary, I. J. (2005). Goldberg's Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, 39, 317-330.
- Kolb, D.A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, N.J.: Prentice-Hall.
- Kolb, A.Y. & Kolb, D.A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning & Education*, 4, 193-212.

- Loo, R. (2002) A meta-analytic examination of Kolb's learning style preferences among business majors. *Journal of Education for Business*, 77, 252-256.
- Marsh, H. W. & Yeung, A. S. (1996). The distinctiveness of affects in specific school subjects: An application of confirmatory factor analysis with the National Educational Longitudinal Study of 1988. *American Educational Research Journal*, 33, 665-689.
- Nijhuis J.F.H., Segers M.S.R., & Gijssels, W.H. (2007). The interplay of personality, perceptions of the learning environment and learning strategies: A study amongst Business Students. *Studies in Higher Education*, 32(1), 59-78.
- Rode, J. C., Arthaud-Day, M. L., Mooney, C. H., Near, J. P., Baldwin, T. T., Bommer, W. H., & Rubin, R. S. (2005). Life satisfaction and student performance. *Academy of Management Learning & Education*, 4, 421-433.
- Schau, C.; Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement*, 55, 868-875.
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.
- Shotick, J., & Stephens, P. R., (2006). Gender inequities of self-efficacy on task-specific computer applications in business. *Journal of Education for Business*, 81, 269-273.
- Stover, D., Morris, J. S., Pharr, S., Reyes, M. G., & Byers, C. R. (1997). Breaking down the silos: Attaining an integrated business common core. *American Business Review*, 15, 1-11.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy -- value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68-81.

APPENDIX

This Appendix extends the Appendix of Chapter 5, focussing on the structural equation model extended with course outcomes. To distinguish between attitudes and outcomes, the vector the measurements on attitudes y is supplemented by measurements on outcomes z . A minor extension with respect to the standard notation for SEM's will be introduced in order to preserve consistency of notation in this Appendix. For each subject the measurement model for the endogenous latent factor is written for $j = 1, \dots, 5$ as

$$\begin{pmatrix} y^{(j)} \\ z^{(j)} \end{pmatrix} = \begin{pmatrix} \Lambda_y^{(j)} & 0 \\ 0 & \Lambda_z^{(j)} \end{pmatrix} \begin{pmatrix} \eta_y^{(j)} \\ \eta_z^{(j)} \end{pmatrix} + \begin{pmatrix} \varepsilon_y^{(j)} \\ \varepsilon_z^{(j)} \end{pmatrix} \text{ with } \text{Cov} \begin{pmatrix} \varepsilon_y^{(j)} \\ \varepsilon_z^{(j)} \end{pmatrix} = \begin{pmatrix} \Theta_{\varepsilon_y}^{(j)} & 0 \\ 0 & \Theta_{\varepsilon_z}^{(j)} \end{pmatrix}$$

where $z^{(j)}$ is an n_j -dimensional vector of observed results for subject j ($j = 1, \dots, 5$) and $\eta_z^{(j)}$ is an m_j -dimensional vector of latent factors (outcomes). The specification of the matrix of factor loadings again imposes a factor complexity of 1 and, hence, contains $9 + n_j$ parameters to be estimated. The covariance matrix of the errors of measurement $\varepsilon_y^{(j)}$, $\Theta_{\varepsilon_y}^{(j)}$, is a diagonal matrix but $\Theta_{\varepsilon_z}^{(j)}$ is allowed to have non-diagonal elements unequal to zero. The measurement model for the exogenous latent factors remains the same, whereas the introduction of the additional endogenous latent factors requires specifying the structural equations as

$$\begin{pmatrix} \eta_y^{(j)} \\ \eta_z^{(j)} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ B_{zy}^{(j)} & 0 \end{pmatrix} \begin{pmatrix} \eta_y^{(j)} \\ \eta_z^{(j)} \end{pmatrix} + \begin{pmatrix} \Gamma_y^{(j)} \\ \Gamma_z^{(j)} \end{pmatrix} \xi + \begin{pmatrix} \zeta_y^{(j)} \\ \zeta_z^{(j)} \end{pmatrix} \text{ with} \\ \text{Cov} \begin{pmatrix} \zeta_y^{(j)} \\ \zeta_z^{(j)} \end{pmatrix} = \begin{pmatrix} \Psi_y^{(j)} & 0 \\ 0 & \Psi_z^{(j)} \end{pmatrix}$$

The number of parameters to estimated in $\Lambda_y^{(j)}$, $\Theta_{\varepsilon_y}^{(j)}$, $\Lambda_x^{(j)}$, $\Phi^{(j)}$ and $\Psi_y^{(j)}$ equals 56 and this number is the same for all subjects. The full covariance matrix of the vector of item-parcels $(y^{(1)'} z^{(1)'} x^{(1)'})'$ of dimension $34 + n_j$ are used for estimation. Table A.1 presents the numbers of estimated parameters in the remaining matrices for each subject, along with the resulting degrees of freedom (normalization constraints have been imposed on the diagonal elements of $\Phi^{(j)}$, $\Psi_y^{(j)}$ and $\Psi_z^{(j)}$).

Table A-1. Number of estimated parameters and degrees of freedom per subject

	$\Lambda_z^{(j)}$	$\Theta_{\varepsilon_z}^{(j)}$	$\Theta_{\delta}^{(j)}$	$B_{zy}^{(j)}$	$\Gamma_y^{(j)}$	$\Gamma_z^{(j)}$	$\Psi_z^{(j)}$	Total estimated parameters
Business Statistics	14	18	33	5	9	4	10	149
Finance & Accounting	0	0	28	1	6	1	1	93
Marketing Management	0	0	28	0	5	1	1	91

Since measurement models for outcome variables with single indicators are set up using a pseudo latent variable number (all subjects except for Business Statistics), entries in the columns of $\Lambda_z^{(j)}$ and $\Theta_{\varepsilon_z}^{(j)}$ are equal to 0.

Multi-group analysis, such as comparing models for female and male students, is performed by estimating the same model on separate sub-samples. The investigation of gender-invariance implies constraining parameter estimates of the same parameters in female and male models to be identical, and contrasting the model fit of the constrained models with the model fit of the unconstrained models.

Chapter 7

POSTSCRIPT

Do subject-specific achievement motivations constitute hierarchic, multidimensional and multifaceted constructs? In Chapters 4 to 6 of this thesis different aspects of this research question have been addressed. The aspect of achievement motivations being multidimensional refers to variations in motivations over academic subject domains. Strong evidence has been found for multidimensionality for university students participating in an international business program. Most empirical studies of achievement motivations over subject domains have focused on motivations for a range of broad domains such as math, reading, sports, or music of children and adolescents in middle and high school (Wigfield & Eccles, 2000, 2002). The academic subjects that are part of a business program are much more congruent than the wide range of different subjects in earlier schooling, so the finding of multidimensionality cannot be regarded as a natural extension of multidimensionality of motivations in middle and high school subjects. In addition to investigating multidimensionality at the level of subjects, the existence of multidimensionality at the aggregated level of categories of subjects has also been investigated, applying a simplified, 'pure' versus 'applied', taxonomy of academic subjects (Alexander, 1992, 1997; Burke & Moore, 2003). The pure/applied decomposition generates patterns of achievement motivations that are very different, both with regard to the internal factor structure of motivations, as well as with regard to relationships with personality factors, indicating that such taxonomy is a useful instrument in the study of learning processes in business education.

The second part of the research question refers to the hierarchic nature of subject-specific achievement motivations (Marsh & Yeung, 1996). Although these motivations are different (as evidenced by the

multidimensionality), they also contain commonalities. Motivations for different subjects contain common parts, and do so not only in a pair-wise manner, but also in the stronger case of investigating motivations for a range of subjects at the same time. That is: subject-specific achievement motivations can be decomposed into a generic component, a trait-like property that expresses itself in the approach of any (business) subject, and a second component being truly specific for the subject under study. On the basis of the methodological guidelines applied in this thesis, it has been assumed that the generic component of motivations is causal for the subject-specific component, implying that the latter component is formed in the interaction of the generic component and the characteristics of the academic subject: the hierarchy. The ability not only to prove the existence of both components, but also to calculate the relative size of both components, is an attractive feature of advanced estimation techniques available in structural equation modelling methods that is relevant for contemporary discussions in motivational research. In addition, these modelling outcomes have implications that can be generalised beyond the level of modelling achievement motivations. Contemporary applications in modelling student learning within the 3P framework indicate that student related characteristics, if not hard encoded, should be regarded as part of the process stage of the learning process: in principle, everything is contextualized. These findings are diametrically opposed to those of early applications within the 3P framework that typically absorb student related variables in the presage part, thereby denying the role of context for these factors. The empirical evidence presented in this thesis, based on structural equation modelling, shows that the results of the early and the contemporary research within the 3P framework should be characterized as taking positions that are too extreme.

The third aspect of the research question refers to the multifaceted nature of achievement motivations. Expectancy-value models of achievement motivations hypothesise that at least three different kinds of motivational variables can be distinguished: two expectancy factors, measuring students' beliefs about their own ability and about the perceived task difficulty, and a subjective task value (Wigfield & Eccles, 2000, 2002). In this thesis, evidence is provided that these three constructs can be empirically distinguished in the domain of business subjects. Beyond the traditional composition of the expectancy-value model, it has been shown that the subjective task value construct can be decomposed into an affective component and a value component that behave rather differently.

As a result, our affect-extended expectancy-value model applied in Chapters 4 to 6 contains four empirically distinguishable facets. In the investigations reported in Chapters 2 and 3 two further facets have been demonstrated: interest, and efforts. All in all, the several chapters provide ample evidence that the empirics of achievement motivations based on the expectancy-value model constitute a completely different world than the layman's view of a student being motivated or not.

The three aspects multifacetedness, multidimensionality, and hierarchy, demonstrate strong dependencies. E.g., the disentangling of affective and value related aspects from the more broader construct "task value" proves to be strongly related to the hierarchic nature of motivations: the affective and value-related constructs take opposite positions along the spectrum of general versus subject-specific. Value appears to be the motivation with the strongest generic component, whereas different subjects have little in common when focusing on affect. This provides further motivation to disentangle the two constructs: not only are they distinguishable in a statistical sense, but in addition the two constructs behave very differently in terms of their variation over academic subjects.

The decomposition of achievement motivations into generic and subject-specific components can be phrased in terms of the malleability of achievement motivations. A construct that is demonstrated to be invariable over different contexts, is by definition much more difficult to change in a favourable direction than a construct that depends on the context. Since education is all about change, this has important practical implications for the type of changes teachers or other elements of the learning & teaching context can bring about. Educational interventions have little perspective if their point of application addresses the least malleable parts of learner's characteristics. Statistical decompositions of learning related constructs might not be illustrative enough to indicate that some interventions are inclined to be to no avail. To that reason, gender and personality traits have been introduced as determinants of subject-specific achievement motivations. To the extent that motivations are explained by these determinants, as demonstrated in Chapters 5 and 6, changing achievement motivations will prove to be ineffective since individual difference theory has shown that these two determinants are among the most stable personality characteristics.

In the first two chapters, the major focus is on the explanation of the existence of naïve statistical theories, measured by levels of statistical

reasoning. The most basic formulation of the research question relevant for this part, is whether ‘naïve knowledge’ can be explained by the same type of factors as ‘academic knowledge’. Models of student learning explain learning achievement by three broad determinants: prior knowledge, strategic processing and motivational and affective factors (Alexander, 2006; Dai & Sternberg, 2004). Chapter 2 of this thesis shows that prior knowledge, interpreted as the level of academic knowledge, is remarkably unrelated to levels of naïve knowledge. This finding corroborates findings of other educational research on naïve theories (Alexander, 2006; Garfield, 2003). So if any explanation of naïve knowledge is to be found, it should be found in the other two factors: strategic processing and motivational and affective factors. Part of Chapter 2 is devoted to the role of strategic processing; the conclusion is that preferences for a reproduction oriented learning pattern help explain the presence of naïve knowledge. In Chapters 2 and 3 the main focus is on the contribution of motivational factors in the explanation of naïve knowledge. Motivational and affective factors are operationalized by a version of the expectancy-value model that contains six constructs: beyond the four constructs discussed in the second part of the thesis, two constructs are added: effort in learning, and interest. Of these six motivational constructs, effort in learning appears to be the strongest predictor of naïve knowledge: students, who express higher levels of learning effort, possess higher levels of misconceptions. Perceived task difficulty has a very modest impact: the more difficult students regard statistical tasks, the higher levels of misconceptions. Both these motivational and strategic processing constructs are gendered phenomena, and thus help explain the clear gender effects found in levels of naïve knowledge.

REFERENCES

- Alexander, P. A. (1992). Domain knowledge: Evolving themes and emerging concerns. *Educational Psychologist*, 27, 33-51.
- Alexander, P. A. (1997). Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 213-250). London, JAI Press.
- Alexander, P. A. (2006). *Psychology in learning and instruction*. Upper Saddle River, NJ: Pearson.

- Burke, L. A., & Moore, J. E. (2003). A perennial dilemma in OB education: Engaging the traditional student. *Academy of Management Learning & Education*, 2, 37-52.
- Dai, D. Y., & Sternberg, R. J. (2004). Beyond cognitivism: Toward an integrated understanding of intellectual functioning and development. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual development and functioning* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum.
- Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1), 22-38.
[Online: [http://www.stat.auckland.ac.nz/%7Eiase/serj/SERJ2\(1\).pdf](http://www.stat.auckland.ac.nz/%7Eiase/serj/SERJ2(1).pdf)].
- Marsh, H. W. & Yeung, A. S. (1996). The distinctiveness of affects in specific school subjects: An application of confirmatory factor analysis with the National Educational Longitudinal Study of 1988. *American Educational Research Journal*, 33, 665-689.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy -- value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68-81.

SAMENVATTING (DUTCH)

Leerprocessen zijn soms uiterst ondoorgrondelijk. Het eerste hoofdstuk van dit proefschrift geeft daar een treffend voorbeeld van. Het vertelt hoe leerlingen, in hun studie van de mechanica, het Newtoniaanse systeem kunnen leren, begrijpen en toepassen binnen schoolse situaties, zoals een toets, maar dat diezelfde leerlingen, geconfronteerd buiten de schoolse context met het probleem van een opgeworpen muntstuk, in de oplossing van dat probleem teruggrijpen op naïeve, ongeleerde en veelal foutieve noties, daarbij hun formele kennis ongebruikt latend. Ook in de statistiek is casuïstiek bekend waar studenten de neiging hebben naïeve concepten, of naïeve redeneervormen, te gebruiken in plaats van eerder geleerde, wetenschappelijke concepten of redeneervormen. De weinige studies die hebben plaatsgevonden naar deze naïeve vormen van statistische kennis geven aan dat het niveau van naïeve kennis (de misconcepties) ongerelateerd is met het niveau van formele kennis (de correcte concepties). Dus naïeve kennis is niet zo zeer complementair aan formele kennis, hetgeen ook onderstreept wordt door het muntvoorbeeld: studenten gebruiken soms naïeve kennis, soms formele kennis, afhankelijk van de context van de probleemstelling. Maar als we naïeve kennis niet kunnen verklaren als een gebrek aan formele kennis, wat zijn dan wel determinanten van die naïeve kennis?

Dit proefschrift stelt zich ten doel bijdragen te leveren aan het doorgronden van leerprocessen, zowel in situaties die ogenschijnlijk zo ondoorgrondelijk zijn als het leren, of ten onrechte niet afleren, van naïeve kennis, als in situaties die al iets beter gekend zijn: het leren van formele kennis. Dit gebeurt vanuit een specifiek methodologisch perspectief: in alle bijdragen worden modellen van leerprocessen geanalyseerd waarin het leerproces wordt verklaard uit zelfopvattingen of zelfconstructies ('self-construals') van studenten. Deze inperking heeft zowel een praktische als een theoretische dimensie. In relatie tot het laatste: in hedendaagse

theorieën over leerprocessen wordt, naast meer objectieve factoren zoals het voorkennisniveau, steeds meer ruimte ingebouwd voor subjectieve percepties van de lerende, als determinanten van leeruitkomsten. Modellen die zich baseren op de betekenis van zelfvertrouwen en eigendunk ('self-concept', 'self-efficacy') voor zowel de manier van leren als de intensiteit van leren, zijn daar sprekende voorbeelden van. De praktische reden om het onderzoek te richten op modellen met zelfpercepties als verklarende grootheden is te vinden in de context waarbinnen het onderzoek heeft plaatsgevonden. Die context is het klaslokaal, en niet het laboratorium, hetgeen het onderzoek 'ecologische validiteit' verschaft, maar tegelijkertijd beperkingen stelt aan de mate waarin er geëxperimenteerd kan worden. Het voordeel van zelfpercepties als belangrijkste informatiebron is dat het vergaren van die gegevens relatief weinig tijd kost, en er om die reden geen beperkingen zijn om met grote groepen van studenten te werken. Sterker nog: door de keuze voor een bepaald type model, dat met latente variabelen, en de omstandigheid dat schattingstechnieken voor modellen met latente variabelen erg data-intensief zijn, was het een buitenkans om het onderzoek te kunnen baseren op verschillende grote cohorten van eerstejaars studenten. De hier gerapporteerde studies zijn daarom niet enkel van belang als bijdrage aan de theorievorming over leerprocessen, maar ook als methodologische bijdragen aan de toepassing van nieuwe technieken op het gebied van latente variabelen modellen, zoals imputatietechnieken, het bundelen ('parcelen') van item-data, tweede-orde confirmatorische factoranalyse modellen, en structurele vergelijkingen modellen.

De vraag wat het niveau van naïeve statistische kennis bepaalt, staat centraal in hoofdstukken twee en drie. Naïeve kennis wordt opgedaan in niet-schoolse contexten, voorafgaand aan de overdracht van formele kennis. Zowel voor statistische concepten als het Newtoniaanse systeem geldt dat de eerste kennismaking met begrippen als kracht en impuls en hun betekenis voor het dagelijkse leven buiten de schoolse context plaatsvindt. Wanneer we de vraag stellen welke student-gerelateerde factoren het niveau van naïeve kennis bepalen, heeft die vraag dus niet primair betrekking op het aanleren van naïeve concepten, maar vooral op het al dan niet afleren ervan. Waarom verdrijft formele kennis, onderwezen lang nadat de naïeve kennis zich heeft gevormd, bij sommigen wel die naïeve kennis, terwijl bij anderen die formele kennis niet voldoende bekijft, en studenten, althans in sommige contexten, terugvallen op de oudere vormen van naïeve kennis? De belangrijkste conclusie van hoofdstukken

twee en drie is dat naïeve kennis op z'n hoogst uiterst zwak met variabelen zoals niveaus van formele kennis en verschillende student-kenmerken is gerelateerd. De zwakke relaties die er wel zijn, vertonen een consistent beeld. Studenten met een leerbenadering die relatief sterk op diepteverwerking is gericht, zijn beter in staat naïeve kennis in te wisselen voor formele kennis, dan studenten met een relatief sterke stapsgewijze (oppervlakte) verwerking. Dat verschil weerspiegelt zich in de gepercipieerde inzet voor de studie. Studenten die aangeven met veel inzet te gaan studeren, scoren hoger op het naïeve kennisniveau, als gevolg van een achterliggende relatie tussen studie-inzet en stapsgewijs leren. En tenslotte is er de relatie met de verschillende typen cursusuitkomsten. De cursus waarbinnen het onderzoek is verricht, het vak kwantitatieve methoden, kent een portfolio van uiteenlopende toetsinstrumenten, met de bedoeling zowel leeruitkomsten die sterker afhankelijk zijn van 'transpiratie', als die welke meer afhankelijk zijn van 'inspiratie' een plaats te geven in het eindoordeel (zoals de beoordeling van huiswerkopdrachten, resultaten op kleine tussentijdse toetsen, de zogenaamde quizzes, en de resultaten op het afsluitende schriftelijke examen). Deze deelresultaten zijn verschillend gerelateerd met gepercipieerde inzet, met als uitersten een positieve relatie met huiswerkscores, en een negatieve relatie met de scores op de schriftelijke toets. Eenzelfde verband is er met het niveau van naïeve kennis: dat is positief gerelateerd aan de huiswerkscore, en negatief gerelateerd aan de toetsscore. Dit mag als een opmerkelijk resultaat worden aangemerkt, wanneer men zich realiseert dat alle prestatiescores gemeten in de portfolio een zeer sterke positieve correlatie kennen. Statistisch redeneren is een heel bijzondere categorie van statistische kennis, waarvan de vorming, evenals het niet afleren van naïeve vormen ervan, nog slecht door leermodellen verklaard kan worden.

Het onderwerp statistisch redeneren, en instrumenten ontwikkeld om de competenties van studenten hierin te meten, zoals de besproken SRA, heeft in de internationale literatuur vooral bekendheid gekregen door het bestaan van 'puzzels'. De eerste puzzel is hiervoor al benoemd en heeft betrekking op het geconstateerde ontbreken van een relatie tussen het vermogen statistisch te redeneren en veel gebruikte maatstaven van cursusprestaties, zoals de scores op toetsen. De tweede puzzel heeft betrekking op landeneffecten: studenten uit verschillende landen presteren zeer verschillend op testen die statistisch redeneren meten. Ook in ons onderzoek vinden we die verschillen, maar tegelijkertijd lijken die niet moeilijk te verklaren, en dus bezwaarlijk als echte puzzels gekarakteriseerd

te kunnen worden. Buitenlandse instromers scoren minder goed op statistisch redeneren dan Nederlandse studenten. Maar dat rijmt uitstekend met het feit dat in het Nederlandse secundaire onderwijs statistiek en kansrekening een betrekkelijk unieke positie innemen, zodat een voor de hand liggende verklaring voor het verschil in statistisch redeneren is dat die voortkomt uit een verschil in scholing. Zo'n evidente verklaring is er allerm minst voor de laatste puzzel. Dat is de puzzel van het genderverschil: vrouwelijke studenten doen het op het gebied van statistisch redeneren slechter dan manlijke studenten, ook als rekening wordt gehouden met verschillen in opleiding of kennisniveau. Dit gendereffect komt overeen met de uitkomsten van internationaal onderzoek: ook daar worden verschillen geconstateerd ten gunste van manlijke studenten. Een deel van de puzzel kan met gebruik van de analyses beschreven in hoofdstukken twee en drie uit achterliggende oorzaken worden verklaard. Vrouwelijke studenten scoren namelijk hoger dan manlijke studenten op de stapsgewijze leerbenadering, en op 'inzetvol leren', en juist die twee variabelen zijn negatief gerelateerd met het niveau van statistische redeneren. Overigens is deze verklaring niet volledig: nog steeds resteert er een, zij het in omvang geringere, genderpuzzel.

Leerinzet, zo cruciaal voor statistisch redeneren, is een factor die tot de prestatiemotivaties wordt gerekend of, wanneer we de terminologie van de wiskundige leertheorieën volgen, met leerattitude wordt aangeduid. In de hoofdstukken vier, vijf en zes verdiepen we ons verder in die prestatiemotivaties, gebruikmakend van een theorie die in veel empirische studies naar individuele verschillen in leren wordt gebruikt: het verwachting*waarde model. De keuze om, en ook hoe hard, te gaan leren wordt in dit model gezien als de uitkomst van een rationeel keuzeprocess. Zoals de naam van het model al aanduidt, wordt de uitverkoren keuze enerzijds bepaald door wat de verwachte uitkomsten van ieder van de keuzeopties zijn, anderzijds door de waarde die aan de verwezenlijking van die uitkomst wordt toegekend. Die verwachting kan verder verbijzonderd worden tot een tweetal determinanten: hoe goed de student zichzelf in de uit te voeren taak acht (de inschatting van eigen competenties) en de gepercipieerde moeilijkheidsgraad van de taak. Ook de waarde van de leeruitkomst valt, in het model dat we voor onze studies hebben gebruikt, uiteen in een tweetal determinanten. De eerste wordt kortweg met waarde aangeduid, en slaat op de meer extrinsieke gerichte opbrengsten van het behalen van de leertaak. De tweede, met affectie aangeduid, heeft primair

betrekking op de intrinsieke aspecten van het succesvol afronden van de leertaak. Het hierboven geschetste model van motivatieprestaties omvat dus in totaal vier verschillende facetten: gepercipieerde competentie, geschatte moeilijkheidsgraad, gepercipieerde waarde van slagen in de taak en affectie.

Dat op een betekenisvolle wijze verschillende facetten onderscheiden kunnen worden in motivationele factoren die het leerproces beïnvloeden is niet vanzelfsprekend. Er bestaan een groot aantal studies die stellen dat, alhoewel vanuit theoretisch perspectief het onderscheiden van facetten mogelijk is, dit vanuit empirisch perspectief niet echt zinvol is. In die studies blijken studenten die hoog scoren op het ene facet, stevast ook hoog te scoren op het ander facet (of facetten). De vraag of in prestatie-motivaties verschillende facetten zijn te onderscheiden, is dus vanuit empirisch perspectief relevant.

Deze vraag is uit te breiden in een tweede richting: zijn prestatie-motivaties multidimensionaal in de zin dat studenten in hun leermotivatie onderscheid maken tussen verschillende type leertaken? Daarbij kan gedacht worden aan de rol van het vakgebied of domein: is de motivatie om een wiskundetaak af te ronden, te vergelijken met die om een marketingvraagstuk op te lossen? Vanuit een theoretisch perspectief lijkt die vraag triviaal: het zijn geheel verschillende leertaken, dus waarom zou de motivatie van een student voor beide gelijk zijn? Tegelijkertijd suggereert empirische literatuur dat het antwoord op die vraag genuanceerder ligt: ook hier kunnen we theoretisch verschillend veronderstelde fenomenen veelal empirisch niet onderscheiden. In andere woorden, een student is vaak ofwel gemotiveerd en is dat dan voor veel verschillende vakken, of is niet, dan wel minder gemotiveerd, wederom ongeacht het vak. Zowel de vraag van de multi-dimensionaliteit als de vraag van het onderscheidbaar zijn van facetten worden in deze studie in een empirische context geanalyseerd. Vanuit een onderzoeksoogpunt bezien uitdagend is deze context uitdagend: richten veel toegepaste studies zich op sterk uiteenlopende vakgebieden als wiskunde en taalonderwijs op middelbare school niveau, onze studies richten zich op een betrekkelijk homogene groep van vakken uit het bedrijfskunde programma. Desondanks bevestigen onze analyses de hypothesen dat prestatie-motivaties te onderscheiden zijn naar facet en vakgebied. Dit is overigens niet het eindpunt van de analyse. Juist omdat de conclusie is dat motivaties variëren, komt er een vervolgvraag naar boven: kunnen die motivaties (en daarbinnen: de verschillende facetten waaruit motivaties zijn

opgebouwd) gesplitst worden in onderliggende componenten, eentje die generiek is, en eentje die vakspecifiek is?

Het beantwoorden van al dit type vragen vergt de inzet van zware statistiek. Wanneer we het niet direct waarneembaar zijn van motivatieconstructies namelijk recht willen doen, zijn latente variabelenmethoden aangewezen, en impliceert de decompositie van prestatiemotivaties in een generieke en een vakspecifieke component bijvoorbeeld het toepassen van tweede-orde confirmatorische factoranalyse. Gebruik makend van de geëigende methodologie leidt dit tot een aantal waardevolle inzichten. Zo blijkt bijvoorbeeld dat daar waar affectie en waarde alle twee betrekking hebben op het waarderingsaspect van het leerkeuze probleem, de twee constructen elkaars tegenpool zijn in het spectrum van generieke versus specifieke motivaties. De intrinsiek georiënteerde factor affectie heeft de kleinste generieke component van alle motivaties, en varieert dus in relatieve zin het sterkst van vak tot vak, terwijl de extrinsiek georiënteerde factor waarde juist de factor is met de grootste generieke component, en dus relatief gezien weinig varieert tussen vakgebieden. Deze uitkomst heeft belangrijke implicaties voor de vormgeving van het curriculum. Daar waar het stimuleren van positieve leermotivaties algemeen gerekend wordt tot één van de doelstellingen van het onderwijs, is het van belang te weten op welk niveau die doelstelling het best nagestreefd kan worden: in de context van het onderwijs van een specifiek vak of mogelijk daarbuiten, zoals in de context van een algemene studievoordigheidstraining. Tevens is het van belang te weten welke rol de verschillende motivationele factoren hebben op de uitkomsten van het leerproces. De uitkomsten van de verschillende structurele vergelijkingen modellen geven daar inzicht in. Leren we uit de factoranalytische studies dat het stimuleren van affectie bij voorkeur in een vakspecifieke context dient te geschieden, aangezien deze factor bij uitstek een belangrijke vakspecifieke component bezit, de structurele vergelijkingen studies wijzen erop dat zelfs binnen deze context het verbeteren van affecties niet het meest effectieve onderwijskundig instrument zal zijn, aangezien de rol van affectie op de uitkomsten van het leerproces gedomineerd wordt door de rol van bijvoorbeeld gepercipieerde competentie. Dit suggereert beperkingen in de effectiviteit van allerlei pogingen om bijvoorbeeld middelbare scholieren te overtuigen dat B-vakken interessant zijn; beter lijkt het ze meer vertrouwen in eigen kunnen mee te geven.

OVER DE AUTEUR (DUTCH)

De auteur is 22 juni 1956 geboren in Delft. Een middelbare schoolopleiding van 1968 tot 1974 aan het Thorbecke Lyceum in Arnhem werd vervolgd met een studie econometrie in Groningen. Gedurende die studie, cum laude afgerond in 1981, was de eerste baan die van student-assistent verantwoordelijk voor de computerberekeningen nodig om de jaarlijkse Grecon-voorspellingen te genereren, en het geven van onderwijs in computerapplicaties aan econometriestudenten. De tweede baan was daar een logisch vervolg op: die van onderzoeksassistent in dienst van ZWO, voorloper van NWO, met als taak het uitvoeren van een onderzoek naar toepassingen van dynamische factoranalyse in het modelleren van econometrische tijdreeksen. Een carrière die in een voortijdig stadium werd ingeruild voor de grote Maastrichtse opdracht: het vanaf begin 1984 helpen vormgeven van de eerste economie opleiding ontworpen volgens de principes van probleemgestuurd onderwijs.

Mijn fascinatie voor onderwijskundige probleemstellingen heeft tevens een belangrijk stempel gedrukt op mijn onderzoeksactiviteiten, zowel in termen van prioriteitsstelling (altijd ondergeschikt aan onderwijstaken), als in termen van onderwerpskeuze. De focus van een puur econometrische benadering middels methoden van dynamische factoranalyse, verschoof naar een meer psychometrische focus, gebruik makend van methoden voor niet-dynamische factoranalyse, toegepast op modellen van leerprocessen. Toch heeft het econometrische onderzoek wel een publicatie opgeleverd, noemenswaardig omdat het in mijn ogen misschien wel de mooiste van m'n publicaties is, zeker gezien het eervolle gegeven dat auteurs als Arnold Zellner en Herbert Simon zich onder de medeauteurs van het boek bevinden:

- Tempelaar, Dirk (2001). Simplicity in a behavioural, non-parametric context. In: Arnold Zellner, Hugo A. Keuzenkamp and Michael McAleer (Eds.),

Simplicity, inference and modelling: keeping it sophisticatedly simple (pp. 227-241). Cambridge: Cambridge University Press.

Zoals aangegeven, het ontwerpen van een probleemgestuurd programma voor kwantitatieve vakken in de opleidingen economie en bedrijfskunde was de hoofdtaak in de Maastrichtse begintijd. Het uitdenken van een bij dat curriculum passend toetsstelsel volgde daar onmiddellijk op. De onderzoeksbijdragen, vooral in EDINEB-verband, getuigen daarvan.

- Tempelaar, Dirk & Schenk, Eugene (1995). Quantitative methods in a problem-based curriculum. In Wim H. Gijselaers et al. (Eds.), *Educational innovation in economics and business administration: The case of problem-based learning* (pp. 203-211). Dordrecht: Kluwer Academic Publishers.
- Tempelaar, Dirk T. (1995). Student assessment in a problem-based curriculum. In Wim H. Gijselaers et al. (Eds.), *Educational innovation in economics and business administration: The case of problem-based learning* (pp. 340-346). Dordrecht: Kluwer Academic Publishers.
- Tempelaar, Dirk T., Wiedersheim-Paul, Finn & Gunnarsson, Elving (Eds.) (1998). *Educational innovation in economics and business II: In search of quality*. Dordrecht: Kluwer Academic Publishers.
- Tempelaar, Dirk T. (1998). Congruence of assessment and instructional system: The case of problem-based learning. In Dirk T. Tempelaar, Finn Wiedersheim-Paul & Elving Gunnarsson (Eds.), *Educational innovation in economics and business II: In search of quality* (pp. 197-212). Dordrecht: Kluwer Academic Publishers.
- Tempelaar, Dirk T. (1998). Selected-response examinations in a student-centred curriculum. In Dirk T. Tempelaar, Finn Wiedersheim-Paul & Elving Gunnarsson (Eds.), *Educational innovation in economics and business II: In search of quality* (pp. 239-254). Dordrecht: Kluwer Academic Publishers.
- Tempelaar, Dirk T. (1998). Computer-based education in a student-centred curriculum. In Richard G. Milter, John E. Stinson & Wim H. Gijselaers (Eds.), *Educational Innovation in Economics and Business III: Innovative Practices in Business Education* (pp. 303-313). Dordrecht: Kluwer Academic Publishers.
- Tempelaar, Dirk T. (1999). Problem-based learning: New ideas from cognitive psychology about teaching & learning. In *Educational change and modern teaching methods* (pp. 221-225). Vilnius: Vilnius Pedagogical University.

- Tempelaar, Dirk T. (1999). The Reform in Teaching Statistics & Mathematics. In *Educational Change and Modern Teaching Methods* (pp. 226-232). Vilnius: Vilnius Pedagogical University.
- Žilinskas, Antanas, Calvin, James, & Tempelaar, Dirk T. (1999). Problem Based Learning and Internet Technology for Studying Courses on Algorithms. In *Educational Change and Modern Teaching Methods* (pp. 274-278). Vilnius: Vilnius Pedagogical University.

Onderzoek naar het ontwerp van programma's en toetsen verschoof in de tijd geleidelijk naar onderzoek over de werking van die programma's: hoe verlopen leerprocessen in die programma's, welke rol spelen individuele verschillen erin, wat doen de verschillende vormen van voorkennis ertoe. Alle bijdragen gebundeld in dit proefschrift vallen in die categorie; niet hierin opgenomen zijn daarenboven een aantal studies naar de rol van metacognitie en van epistemologische opvattingen.

- Tempelaar, Dirk T. (2002). Modelling students' learning of introductory statistics. In *Proceedings 6th International Conference on Teaching Statistics*. Cape Town: IASE.
- Tempelaar, Dirk T. (2003). Statistical reasoning and its relationship to attitudes toward statistics and achievement. In *Proceedings 54th ISI session*. Berlin: ISI.
- Tempelaar, Dirk T. (2004a). Statistical Reasoning Assessment: an Analysis of the SRA Instrument. In *Proceedings of the ARTIST Roundtable Conference on Assessment in Statistics*.
- Tempelaar, Dirk T. (2004b). Statistical Reasoning Assessment: an Analysis of the SRA Instrument. In *ASA Proceedings of the Joint Statistical Meetings 2004*, pp. 2797-2804, Alexandria, VA: American Statistical Association.
- Tempelaar, Dirk T. (2005). Creating conditions for collaborative learning. In Richard G. Milter, Valerie S. Perotti, & Mien S. R. Segers (Eds.), *Educational innovation in economics and business IX: Breaking boundaries for global learning* (pp. 239-254). Dordrecht: Kluwer Academic Publishers.
- Tempelaar, Dirk T. (2006a). The role of metacognition in business education. *Industry & Higher Education*, 20 (5), 291-298.
- Tempelaar, Dirk T. (2006b). A structural equation model analyzing the relationship between students' statistical reasoning abilities, their attitudes toward statistics, and learning approaches In Allan Rossman & Beth Chance (Eds.), *Proceedings 7th International Conference on Teaching Statistics*. San Salvador: IASE.

- Tempelaar, Dirk T., Gijselaers, Wim H. & Schim van der Loeff, Sybrand (2006). Puzzles in Statistical Reasoning. *Journal of Statistics Education*, 14 (1).
- Tempelaar, Dirk T., Schim van der Loeff, Sybrand & Gijselaers, Wim H. (2006). Students' self-theories, goal orientations and its relationship to expectancy-value based achievement motivations. In *Proceedings of the 4th International Biennial SELF Research Conference*. Michigan: SELF.
- Tempelaar, Dirk T., Gijselaers, Wim H., Schim van der Loeff, Sybrand & Nijhuis, Jan F. H. (2007). A structural equation model analyzing the relationship of student achievement motivations and personality factors in a range of academic subject-matter areas. *Contemporary Educational Psychology*, 32 (1), 105-131.
- Tempelaar, Dirk T., Schim van der Loeff, Sybrand & Gijselaers, Wim H. (2007). Students' self-theories, goal orientations and its relationship to expectancy-value based achievement motivations. In *Proceedings of the 4th International Biennial SELF Research Conference*. Sydney: SELF.
- Tempelaar, Dirk T. & Nijhuis, Jan F. H. (2007). Commonalities in attitudes and beliefs toward different academic subjects. In M. K. McCuddy, H. van den Bosch, J. W. B. Martz, A. V. Matveev & K. O. Morse (Eds.), *Educational innovation in economics and business X: The challenges of educating people to lead in a challenging world* (pp. 225-250). Berlin: Springer.
- Tempelaar, Dirk T., Schim van der Loeff, Sybrand, & Gijselaers, Wim H. (2007). A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities and course performance, manuscript in revisie voor publicatie in *Statistics Education Research Journal*.
- Tempelaar, Dirk T., Schim van der Loeff, Sybrand, & Gijselaers, Wim H. (2007). Variations in achievement motivations over business subjects. Manuscript ter publicatie aangeboden.

Het meest recente onderzoek, eveneens niet in dit proefschrift opgenomen, komt voort uit de Maastrichtse participatie in landelijke bijspijskerprojecten. Samen met vooral Bart Rienties hebben we de resultaten van opeenvolgende 'WebSpijkeren' projecten in de volgende publicaties naar buiten gebracht.

- Rienties, Bart, Dijkstra, Joost, Rehm, Martin, Tempelaar, Dirk T. & Blok, Geke (2005). Online bijspijskeronderwijs in de praktijk. *Tijdschrift voor hoger onderwijs* 23 (4), 239-253.

- Rienties, Bart, Tempelaar, Dirk T., Waterval, Dominique, Rehm, Martin & Gijselaers, Wim H. (2006). Remedial online teaching on a summer course, *Industry & Higher Education*, 20 (5), 327-336.
- Tempelaar, Dirk T., Rienties, Bart & Gijselaers, Wim H., (2006). Internationalisering; en de Nederlandse student? Maastrichtse ervaringen met Nederlandse en Duitse studenten. *Onderzoek van Onderwijs*, 35 (3), 40-45.
- Tempelaar, Dirk T., Rienties, Bart, Rehm, Martin, Dijkstra, Joost, Arts, Mark & Blok, Geke (2006). An online summercourse for prospective international students to remediate deficiencies in math prior knowledge: The case of ALEKS. In *Proceedings WebALT2006* (pp. 23-36). Technical University of Eindhoven: Oy WebALT Inc.
- Tempelaar, Dirk T., Rienties, Bart & Gijselaers Wim H., (2007). Internationalisering; en de Nederlandse student? Deel 2: Leerbenaderingen. *Onderzoek van Onderwijs*, 36 (1).
- Tempelaar, Dirk T., (2007). Onderwijzen of bijspijkeren? *NAW*, 5/8 (1), 55-59.