Change in learning strategies during and after the transition to higher education: The impact of statistical choices on growth trend estimates

Liesje Coertjens

Promotors: prof. dr. Vincent Donche
prof. dr. Peter van Petegem

Dissertation for the degree of doctor in Educational Sciences
at the University of Antwerp
Antwerp, 2014

Universiteit Antwerpen

Universiteit Antwerpen

Institute of Education and
Information Sciences

Change in learning strategies during and after the transition to higher education: The impact of statistical choices on growth trend estimates

Dissertation for the degree of doctor in Educational Sciences

at the University of Antwerp

to be defended by

Liesje COERTJENS

October 13th, 2014

Promotors

prof. dr. Vincent Donche

prof. dr. Peter van Petegem

Doctoral Commission

Prof. dr. Vincent Donche, University of Antwerp
Prof. dr. Sven de Maeyer *(chair)*, University of Antwerp
Prof. dr. Peter van Petegem, University of Antwerp
Prof. dr. John T.E. Richardson, The Open University, UK

Composition of the Doctoral Jury

Prof. dr. Vincent Donche, University of Antwerp
Prof. dr. Joop Hox, Utrecht University
Prof. dr. Sven de Maeyer, University of Antwerp
Prof. dr. John T.E. Richardson, The Open University, UK
Prof. dr. Elke Struyf *(chair)*, University of Antwerp

Cover

design & typesetting: Nik Römer

ISBN

Universitas, Antwerpen

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ASI	Approaches to Study Inventory
ASSIST	Approaches to Study Skills Inventory for Students
CI	confidence interval
CFI	Comparative Fit Index
EPC	expected parameter change
GPA	grade point average
HE	higher education
H&G	Hedeker & Gibbons
ILS	Inventory of Learning Styles
ILS-SV	Inventory of Learning Styles – Short Version
LD	listwise deletion
LG	latent growth
LMI	longitudinal measurement invariance
MAR	missing at random
MCAR	missing completely at random
MI	multiple imputation
MIaux	multiple imputation with auxiliary variables
MILG	multi-indicator latent growth model
ML	maximum likelihood
MLaux	maximum likelihood with auxiliary variables
MNAR	missing not at random
Mod.Ind.	modification indices
PM	pattern mixture
RI	random intercepts
RS	random slopes
RQ	research question
SAL	students' approaches to learning
SE	secondary education
SPQ	Study Process Questionnaire
T/LE	teaching/learning environment
WLSMV	weighted least squares means-variance

Acknowledgements

Writing this dissertation was not a walk in the park. Rather, it was a marathon. Fortunately, a lot of people ran alongside me during a part, or encouraged me from the sidelines. I was so fortunate to have the help of three counselors, Vincent, Peter and Sven. I want to express my thanks to you all for creating and safeguarding a position in which writing the dissertation was do-able, and for the always open door to your offices. Moreover, during our council meetings, you three pointed out the possible directions I could take, but left it up to me to decide which direction to go in. As such, this feels very much 'my own work', although knowing that it would have looked a great deal different without you three running alongside.

John, you followed my work from a not much greater distance, and later completed the membership of the guidance committee. Thanks for your the detailed consideration of my writings and for sharing parts of your - what appears to be - unlimited knowledge base.

I am very happy to have been working with a group of colleagues who celebrated when things were going well and who did not hesitate to help out when they were not. Also, the 'there’s more to life than a PhD' mentality I really appreciated. A special thanks to Alexia, Heleen and Tine for their efforts in collecting the five wave sample used in this work, and to Chris for the excellent job of guiding me through the final practicalities and in organising this defence.

While finalising this PhD, a new project in which I took part had already lifted off. Needless to say this was a rather challenging combination. I want to thank the D-PAC colleagues for cutting me some slack. Moreover, Marije, Roos and Tine, thank you for taking up a large part of the to-do list.

Having finished the marathon, I am convinced that behind every young parent working towards a PhD, there is a solid partner and a whole set of extra helping hands. Moeke and Vake, Armand and Anne-Mie, taking Alex for long walks during my maternity leave so that I could work on a major revisions, the countless shirts ironed, helping with the surprise painting and working on the house when I was at a conference, watching Alex and Ruben particularly often during the final months, this does not even begin to sum up the things you have done to help. You kept on saying you were happy to help, and perhaps the strong connection with your grandchildren that you gained from it played a tiny role in that, but it was a big help to Bene and me, allowing us to really be there for Alex and Ruben during the time left.

Bene, getting this dissertation finished was without doubt the hardest on you. The extra work hours came at the cost of us having little time for the good things in life. My love, thanks for your devotion: not once did you question whether I could do whatever it took, but were all ears when I raised any doubts. Alex and Ruben, for me, you two growing up, being all silly and giggly and discovering the world, put things into perspective and, as such, highlighted what I liked so much about working on this dissertation. Without knowing it, you were my great helpers. To my three pretty-eyed men, I love you.

1. General introduction

"An important question is whether the established developments are specific for this period, or whether they will continue throughout the study. Furthermore, it is important to get information about the developmental pattern. Is it a one-way, gradual process in which students become more self-regulated, deep-level learners? Or is it a capricious pattern, with periods of stability followed by periods of change?"
(Vermetten, Vermunt, & Lodewijks, 1999b, p. 238)

"It is unfortunate that the methods by which longitudinal data are often analyzed are not commensurate with the level of effort involved in their collection"
(Gibbons et al., 1993, p. 739)

One of the main goals of education is preparing students for lifelong learning (Endedijk & Vermunt, 2013; Segers, Nijhuis, & Gijselaers, 2006). From an employment-perspective, two reasons emphasise the importance of this goal. Firstly, education is not able to prepare students for the plethora of challenges they will encounter in the workplace. To tackle these, continuous learning will clearly be required (Marsh, Hau, Artelt, Baumert, & Peschar, 2006). Secondly, the knowledge base that students acquire during education nowadays has a rapid expiry date. Once graduated, they will have to keep their professional expertise up-to-date (e.g., Brooks & Everett, 2008; Shin, Braines, & Johnston, 1993).

To be able to do so, a set of skills for lifelong learning is required. Two important ones are frequently mentioned. Firstly, when the (university) teacher no longer guides the learning, self-regulation skills are essential (Brooks & Everett, 2008; Gow & Kember, 1990; Segers et al., 2006; Vermunt, Richardson, Donche, & Gijbels, 2014). Secondly, critical thinking skills are necessary to form opinions about new content and to solve ill-defined problems (Asikainen, 2014; Gent, Johnston, & Prosser, 1999; Gow & Kember, 1990; Segers et al., 2006; The European Commission, 2010; Vermunt et al., 2014). The use of strategies fostering in-depth understanding is seen as a precondition for these critical thinking skills (Gow & Kember, 1990; Hoeksema, van de Vliert, & Williams, 1997; Reid, Duvall, & Evans, 2005).

Thus, to prepare for lifelong learning, an important challenge for education is to increase students' self-regulated and deep learning. For educational researchers, this raises the question of whether this aim is being met: do students' learning strategies on average evolve in the direction of self-regulated and deep learning? (Segers et al., 2006; Vanthournout, 2011) In addition to the "whether", researchers are also interested in the "how": are the changes in learning strategies to be understood as a gradual trend, or rather as a capricious pattern (Vermetten et al., 1999b)? Moreover, it can be questioned as to whether education meets the aim for all students, or whether some students are falling behind in acquiring deep and self-regulated learning. In this light, examining whether or not the trajectory of change differs between students is also important.

Over the last few years, the number of studies assessing the change in students' learning strategies is increasing (Vanthournout, Donche, Gijbels, & Van Petegem, 2011). Though this can only be applauded, two shortcomings with regard to current practice hamper our understanding of the change in students' learning strategies: (1) studies are limited to the context of higher education and, (2) the set of statistical choices can compromise the accuracy of the estimates of change. Prior to detailing these shortcomings, I will describe the research tradition on learning strategies and elaborate on the findings regarding change in these strategies. After focusing on the shortcomings with regard to current research practice, the focus and overview of the dissertation are presented.

1.1. Research on learning strategies: the SAL tradition

One of the educational theories that aims to develop an understanding of student learning is the Students' Approaches to Learning (SAL) tradition (Richardson, 2011). The founding fathers of this tradition, Marton and Säljö (1976), detected that students differ in how they go about learning and that these differences are related to the quality of learning. This viewpoint on student learning remains up to the present day, the foundation of the SAL tradition.

What has changed is how these differences in learning are captured. Whereas Marton and Säljö (1976) examined students' learning strategies while they were performing a task, the SAL tradition has shifted towards mapping students' general preferences for, or predisposition with regard to, learning (Entwistle & McCune, 2004; Lonka, Olkinuora, & Mäkinen, 2004; Richardson, 2004; Vanthournout, Donche, Gijbels, & Van Petegem, 2014). As noted by Biggs (1993, p. 5), this is "one step removed from what they actually do when engaging a given task in a particular context."

Prior to detailing the two frequently used SAL models, I concur with Entwistle and McCune (2004, p. 339) that one of the difficulties in reading the SAL literature is "the different meanings given to the same term, and the existence of different terms apparently covering the same aspect of studying". For this reason, a glossary is included in the appendix, providing a description for the most frequently used terms in this dissertation.

1.1.1. Two frequently used SAL models

In the SAL tradition, two models are frequently used (Vanthournout et al., 2014). I will first describe how each model views student learning. Subsequently, I will compare how both model conceptualise learning strategies.

First, the *SAL model* describes differences in student learning by means of various approaches, being the combination of a learning strategy (how a student learns) and a motivational component (why a student learns) (Biggs, 1993; Entwistle & McCune, 2004). A deep approach can be conceptualised as using strategies to create meaning (e.g., relating one aspect of the content with another) combined with aiming for understanding. A surface approach can be understood as using memorizing techniques (e.g., learning by heart), with the aim of passing the course or completing the task.

Second, the *learning pattern model* describes differences in student learning in terms of learning patterns (Vermunt & Vermetten, 2004). As shown in Table 1.1., these patterns are a combination of four components. First, processing strategies map the cognitive activities a student habitually applies whilst studying. Second, regulation strategies concern the meta-cognitive activities that students usually rely on to direct their learning process, such as planning or monitoring. These two elements – processing strategies and regulation strategies – are subsumed under the concept of ‘learning strategies’. Third, orientations to learning map students' motives for studying. Fourth, students' conceptions of learning are beliefs about what learning is, for example the absorption of knowledge or the construction of knowledge.

Table 1.1. Overview of the learning patterns of the Vermunt model

		Meaning-oriented	Reproduction-oriented	Undirected	Application directed
Learning strategies	Processing strategies	Deep	Stepwise	Hardly any	Concrete
	Regulation strategies	Self-regulation	External	Lack of regulation	Both external and self-regulated
	Orientations to learning	Construction of knowledge	Intake of knowledge	Cooperation and being stimulated	Use of knowledge
	Conceptions of learning	Personally interested	Certificate or self-test oriented	Ambivalent	Vocation oriented

Source: Based on Vermunt (1996)

From these four components, four superordinate learning patterns were identified (Vermunt & Vermetten, 2004). First, the meaning-oriented learning pattern consists of a combination of deep processing strategies, self-regulation, being personally interested and viewing learning as the construction of knowledge. Second, the reproduction-oriented learning pattern is an integration of stepwise processing (i.e., analyzing and learning by heart), external regulation, viewing learning as the intake of knowledge and being certificate-oriented or wanting to prove oneself (self-test oriented). Third, the undirected learning pattern is characterized by a lack of processing and regulation strategies, an ambivalent learning orientation and seeing learning as a cooperative process, and of being stimulated by the learning environment. Fourth, the application-directed learning pattern consists of concrete processing, being vocation-oriented and viewing learning as using knowledge.

After describing the two most frequently used SAL models, their view on learning strategies can be compared. In the SAL model, learning strategies are described as being either deep or surface, and map how students process the learning content. In the learning pattern model on the other hand, learning strategies are viewed as processing strategies on the one hand, involving deep and stepwise strategies, and regulation strategies on the other hand, which map the steering of the learning process. In sum, while the SAL model views learning strategies as processing strategies, the learning pattern model adds regulatory strategies. In the next section, I will link these learning strategies to the skills described as being beneficial for lifelong learning.

1.1.2. Learning strategies beneficial for lifelong learning skills

From the viewpoint of lifelong learning, self-regulation as well as critical thinking has been discerned as being important skills (see above). Given that self-regulatory skills are not mapped in the SAL model, I opted for the learning strategy scales from the learning pattern model. The self-regulation component captures the degree to which students take their learning process into their own hands.

As shown in Table 1.2., two other regulatory strategies are included in the learning pattern model. External regulation concerns relying on guidance provided by the (university) teacher or learning material. As described by

Vermunt and Vermetten (2004), conceptually, and compared to self-regulation, this regulation strategy is less beneficial from the perspective of lifelong learning. The last regulatory strategy, lack of regulation, captures being undirected in the learning process, and is also viewed conceptually as being less adequate for lifelong learning than self-regulation (Vermunt & Vermetten, 2004). Moreover, given that lack of regulation is associated with the non-completion of higher education studies (Vanthournout, Gijbels, Coertjens, Donche, & Van Petegem, 2012), it may indicate *un*preparedness for lifelong learning.

Table 1.2 The learning strategies from the learning pattern model, their meaning and conceptual link to lifelong learning

Learning strategy		Meaning	Conceptual link to lifelong learning°
Regulation strategies	Self-regulation	Guiding their own learning process	Beneficial
	External regulation	Relying on clues in the learning material or from the teacher to guide the learning process	Less beneficial
	Lack of regulation	Being undirected in the learning process	Less beneficial
Processing strategies	Deep processing		
	Relating and structuring	Relating aspects of the content	Beneficial
	Critical processing	Critically assessing the learning content	Beneficial
	Stepwise processing		
	Memorizing	Learning content by heart	Less beneficial
	Analysing	Examining the learning materials from start to finish	Less beneficial

° Based on Vermunt and Vermetten (2004)

Linked to the second element, critical thinking skills, students' habits with regard to deep learning are considered a precondition for lifelong learning (Gow

& Kember, 1990). In the learning pattern model, students' deep processing strategies are captured using two scales: 'relating and structuring' and 'critical processing' (see Table 1.2.). These deep processing strategies are regarded, at the conceptual level, as being beneficial for lifelong learning (Vermunt & Vermetten, 2004).

Next to the two deep processing strategies, the learning pattern model includes two scales mapping stepwise processing strategies. The memorizing scale maps the degree to which students learn content by heart, while the analysing strategy captures to what extent students process the learning content from start to finish. From the lifelong learning perspective, Vermunt and Vermetten (2004) conceptually judge these stepwise processing strategies to be less adequate than deep processing strategies.

In the learning pattern model, the processing and regulation strategies are related to one another. For example, as shown in Table 1.1., deep processing strategies and self-regulation load on the meaning-oriented learning pattern. When investigating change over time, one has the option between mapping change at the level of the higher order concepts or at the learning strategy level. A possible downside of the first option is that change in one learning strategy can be clouded by stability in another (Vanthournout et al., 2011). For this reason, I argue that researchers interested in change in processing and regulation strategies should map change at the learning strategy level.

We note that in the learning pattern model a fifth strategy, concrete processing, is distinguished. This strategy maps the degree to which students apply the learning content. However, this processing strategy was found to be interwoven with deep processing strategies (Vermunt, 1998; Vermunt & Vermetten, 2004) especially with younger learners and first year students in higher education (Vermunt & Verloop, 1999). Given the expectation of limited added value of including the concrete processing scale in addition to deep processing strategies, we opted to exclude this fifth processing strategy.

Summing up, from the viewpoint of lifelong learning, conceptually, students' self-regulatory and deep processing strategies ideally increase during education, while stepwise processing strategies, external regulation, and, lack of regulation ideally decrease.

1.1.3. The validity of self-report questionnaires

Both the SAL and the learning pattern model rely strongly on self-report questionnaires incorporating Likert-type scales. Respondents of the Study Process Questionnaire (SPQ; Biggs, Kember, & Leung, 2001) score items from (1) "this item is never or only rarely true of me" to (5) "this item is always or almost always true of me". In the Inventory of Learning Styles (ILS) Questionnaire participants are asked to denote how often they apply a certain learning strategy, by scoring it from (1) "I never or hardly ever do this" to (5) "I (almost) always do this".

It has been questioned whether or not the results of these self-reports can be inferred to the construct 'learning strategies'. Possibly, students' recall of learning is inaccurate or social desirability is mapped (Endedijk & Vermunt, 2013; Richardson, 2004). A number of studies have therefore focused on methodological triangulation (Cohen, Manion, & Morrison, 2008). Schatteman and colleagues (1997) for example, used interviews in addition to the ILS. Their finding was that the activities elaborated on in the interviews were consonant with the data from the self-report questionnaire. Endedijk and Vermunt (2013) triangulated the findings from a self-report questionnaire with findings from structured learning reports in which student teachers reported multiple learning activities. Once more, the conclusion was that the results from the structured learning reports were significantly and meaningfully related to the student teachers' preferences for learning as captured by the self-report questionnaire. In sum, this evidence on convergent validity underpins the inference of results from self-report questionnaires to 'the way respondents usually go about learning'. Changes in these self-report scores can therefore be examined to verify whether and how learning strategies change over time.

1.2. Change in learning strategies

Prior to describing the results on the change in learning strategies, I note that in the methodological literature, growth is described as a significant change in a construct, without implying a positive judgment with regard to such a change (Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004; McArdle & Nesselroade, 2013; Singer & Willet, 2003). For example, if respondents significantly increase their alcohol use over time, it is noted that there is growth, clearly without suggesting that this trend should be applauded (e.g., Li, Duncan, Duncan, & Hops, 2001). The meaning of the term 'growth' in the methodological literature thus parallels that of 'change' in the learning strategies' literature. Therefore, in this dissertation, the terms change and growth are used interchangeably.

In term of such growth, two types can be discerned. First, average growth indicates the extent to which and how students evolve on average. Second, differential growth regards the extent to which and how students vary in their change over time. In this dissertation, the term 'growth trend' refers to both types of growth taken together.

To map this growth trend, three waves of data are put forward as a minimum (Singer & Willet, 2003). To my knowledge, in the learning strategies literature, five studies adhere to this criterion (Busato, Prins, Elshout, & Hamaker, 1998; Donche, Coertjens, & Van Petegem, 2010; Severiens, Ten Dam, & Van Hout-Wolters, 2001; Van der Veken, Valcke, De Maeseneer, & Derese, 2009; Vanthournout, 2011). Each of these studies was conducted in the context of higher education. In the next two sections, the findings regarding the growth trend will be considered.

1.2.1. Average growth

Regarding deep processing strategies, students in higher education were found to increase their reliance on these strategies over time (Donche et al., 2010; Severiens et al., 2001; Van der Veken et al., 2009). However, Vanthournout (2011) concluded that only the relating and structuring scale increases, while the degree of critical processing remains constant. With regard to stepwise processing strategies and the subscale analysing, a constant trend was identified (Donche et al., 2010; Severiens et al., 2001; Vanthournout, 2011). The degree of memorizing decreased over time (Donche et al., 2010) or showed a quadratic trend with a rise after an initial drop (Vanthournout, 2011).

Concerning regulation strategies, self-regulation was found to increase (Donche et al., 2010; Severiens et al., 2001; Vanthournout, 2011), or to remain constant (Van der Veken et al., 2009). External regulation, on the other hand, decreased (Donche et al., 2010; Severiens et al., 2001; Vanthournout, 2011). Lastly, undirected learning was found to remain constant (Severiens et al., 2001; Van der Veken et al., 2009) or decrease over time (Donche et al., 2010; Vanthournout, 2011).

In contrast to the other studies, Busato and colleagues (1998) examined change in the higher-order concept of learning patterns. They detected an increase in meaning-oriented learning, which is in line with the findings of the increase with regard to deep processing and self-regulation. The reproduction-oriented learning pattern remained constant. This is at odds with the findings of a decrease in external regulation but in line with the constant trend in stepwise processing strategies. The undirected learning pattern also remained constant, which is in line with findings on the constant trend in terms of lack of regulation.

In sum, the average growth in learning strategies appears to be a move in the direction of deep and self-regulated learning and away from memorizing strategies and externally regulated learning, while the degree of analysing remains constant. The results for lack of regulation appear inconclusive between a constant trend and a decrease.

1.2.2. *Differential growth*

Two of the five studies noted above examined whether students evolve differently over time. Donche et al. (2010) examined whether or not students' changes in learning strategy scales during their time in higher education was dependent on the learning pattern upon entry into higher education. Students exhibiting a meaning-oriented learning pattern in the first year of higher education increased their reliance on self-regulatory strategies to the detriment of external regulation. Their peers with a reproductive/undirected learning pattern increased their use of both deep processing and self-regulatory strategies. Students with a flexible learning pattern, combining deep and surface learning activities in the first year of higher education, increased the critical processing of the learning content to the detriment of memorizing. These students also demonstrated a reduction with regard to external regulation and lack of regulation.

Relying on a multilevel model, Vanthournout (2011) found a differential evolution in change over time for the critical processing, analysing, self-regulation and external regulation scales. For two of these scales, this growth over time was related to the students' initial scores. First, students scoring higher on analysing at the start of higher education tended to decrease their reliance upon it, while those initially scoring lower tended to increase their reliance upon it. Second, in terms of the external regulation scale, the findings suggested that students with a strong preference for external regulation at the start of their higher education decreased their reliance on external sources of regulation at a greater rate. Given that, from the viewpoint of lifelong learning, it is best if external regulation decreases (see 1.1.2), this trend can be judged to be a positive one: students who initially scored higher, decreased faster.

In sum, both studies provide some evidence that during higher education the average growth trend does not appear valid for all students. In Vanthournout's study (2011), for two out of seven scales, the differences between students decreased over time. In the study by Donche and colleagues (2010), students with a reproduction-oriented/undirected learning pattern, scoring initially lowest on deep processing and self-regulation compared to their peers with a meaning-oriented or flexible learning pattern, increased on these scales. Students with a flexible learning pattern, scoring highest on memorizing, external regulation and lack of regulation in the first year, decreased on these scores from the first to the third year of higher education. Thus, though the analytical technique (cluster analysis followed by paired samples *t*-test) does not allow significance testing of the differential growth, the findings of the Donche et al. (2010) study partly point towards a decrease in variability over time as detected by Vanthournout (2011).

1.3. Shortcomings in the current research

An examination of how the studies on the change in learning strategies are undertaken reveals two major shortcomings: the stable educational context and the accuracy of the estimates.

1.3.1. Studies are limited to stable educational contexts

Current research on the change in learning strategies concludes that students' learning strategies are relatively stable (Vanthournout et al., 2011; Vermunt & Vermetten, 2004) and that there is a gradual trend towards deep and self-regulated learning. These studies have all been undertaken in one educational context, that of higher education. As pointed out by Richardson (2011), plausibly, a rather stable educational context gives way to rather stable learning strategies. As such, restricting studies to one educational context can cloud our view on whether the growth in learning strategies is to be understood as a gradual trend, or rather as a capricious pattern, a question asked by Vermetten, Lodewijks and Vermunt (1999b) 15 years ago. Thus, looking at learning strategies when the educational context does differ is warranted.

One important change in terms of educational context is the transition from secondary to higher education. Qualitative research in the United Kingdom has shown that this transition is perceived as a huge culture change, which is accompanied by a 'learning shock': students reported to no longer feeling competent as students (Christie, Tett, Cree, Hounsell, & McCune, 2008). In line with these findings, researchers in the SAL domain have hypothesized that when students enter a new educational context, this can induce a period of friction in which students need to adjust their way of going about learning to the new demands (Lindblom-Ylänne, 2003; Vermunt & Vermetten, 2004). On the other hand, it has been hypothesized that students, when confronted with a changing educational context, initially rely strongly on their usual way of going about learning, rather than changing their learning strategies (Cliff, 2000; Segers et al., 2006).

The two hypotheses above concern students' average growth. Next to this, differential growth is of interest as well: do students vary in their growth over time during the transition from secondary to higher education? As mentioned above, studies examining differential growth are scarce (see 1.2.2.). For a number of scales, the findings point to a decreasing variability between students over time. Does a similar trend show when differential growth is assessed during a change in educational context?

In sum, to investigate hypotheses regarding average growth in a changing educational context, and to enlighten us on differential growth during such a change in educational context, research on the change in learning strategies during the transition from secondary to higher education is called for.

1.3.2. Accurate estimation of the growth trend

In the current practice of modelling growth in learning strategies over time, three elements hamper the accuracy of growth trend estimates: the measurement model underlying the scale scores is neglected, the assumption of longitudinal measurement invariance (LMI) is not tested for, and students with one or more missing data points are removed from the sample.

A. *The measurement model is neglected*

Examining how data on the change in learning strategies are mostly analysed reveals a predominance of manifest scale scores: all five longitudinal studies relied on them (Busato et al., 1998; Donche et al., 2010; Severiens et al., 2001; Van der Veken et al., 2009; Vanthournout, 2011). These manifest scale scores are computed by averaging the scores on the items per scale for each student at each wave. Analysis of change is then conducted on these manifest scale scores.

As such, the multiple items underlying a scale score are neglected when change over time is assessed. Consequently, the measurement error captured when mapping learning strategies is ignored. It is however not debated that learning strategy items measure a certain concept imperfectly: studies using confirmatory factor analyses on questionnaires in the SAL tradition acknowledge that learning strategy items map measurement errors as well (e.g., Biggs, Kember, & Leung, 2001; Boyle, Duffy, & Dunleavy, 2003). Such error can be random or systematic (i.e., capturing another dimension besides the intended construct, Wu, Liu, Gadermann, & Zumbo, 2010). In longitudinal datasets, for the same item, this systematic error is likely to occur at each wave. Therefore, errors pertaining to the same item are often found to correlate over time (Marsh & Hau, 2007; Vaillancourt, Brengden, Boivin, & Tremblay, 2003; Wu et al., 2010).

By computing manifest scale scores, errors and thus also the correlation between errors over time, is ignored. When, prior to estimating growth, random and systematic error is not separated from the true score, or when the correlation between errors remains unaccounted for, estimates of average and differential growth can be inaccurate. Consequently, the validity of the conclusions arrived at from the estimates can be under threat (Shadish, Cook, & Campbell, 2002) in that, possibly, different conclusions might have been reached had these elements been taken into account (Marsh & Hau, 2007). Put differently, results may suggest an absence of growth, while in fact there has been such growth. Other possibilities are that claims on the strength of the growth are invalid or growth can be concluded on whereas, in fact, it is absent.

The methodological literature therefore recommends modelling such error and the correlation between errors explicitly, when estimating growth (Burt & Obradović, 2013; Marsh & Hau, 2007; Shadish et al., 2002). This can be done using a latent growth model which models the multiple indicators underlying a scale score, i.e. a multi-indicator latent growth model (MILG, Muthén & Muthén, 2010; Wu et al., 2010). Studies detailing this MILG, and contrasting the results with those obtained when the measurement model is ignored, are however lacking in the learning strategies domain.

A second downside of neglecting the measurement model underlying scale scores is that measurement invariance cannot be tested for (Marsh & Hau, 2007; Wu et al., 2010). The next section deals with the importance of such LMI testing.

B. Longitudinal measurement invariance is not tested for

A pitfall of measuring a construct at multiple times is that such a measurement can be sensitive to age and/or treatment (Shadish et al., 2002). Students having more experience in studying in higher education could interpret learning strategy items differently from novices. Therefore, prior to modelling growth over time, change due to actual alterations in learning strategies needs to be disentangled from change in the measurement over time (Vaillancourt et al., 2003; van de Schoot, Lugtig, & Hox, 2012; Wu et al., 2010). When manifest scale scores are relied upon without confirmation of the measurement invariance over time, estimates in the growth trend can be partially due to change in the measurement. With this, the validity of the conclusions reached from those estimates regarding the presence, absence or strength of growth can be compromised (Marsh & Hau, 2007; van de Schoot et al., 2012).

Recently, research into changes in learning strategies have increasingly allowed for longer time intervals (Vanthournout et al., 2011). Moreover, to adequately conduct studies on the change in learning strategies when transitioning between educational contexts, such longer time intervals will also be required. In addition, "whereas it may be reasonable to assume the invariance of these properties over short intervals, this assumption becomes more problematic as time intervals become longer" (Marsh & Grayson, 1994, p. 334). Therefore, verifying whether the questionnaire measures equivalently over waves prior to modelling growth is becoming increasingly important in the learning strategies domain.

Though the methodological literature recommends testing for LMI prior to assessing growth, studies on the growth in learning strategies have neglected to do so, perhaps due to a lack of familiarity with the assumption, or with the method of analysis required to verify this. Studies showcasing the measurement invariance testing procedure and the impact of this on growth trend estimates, are therefore called for.

Next to ignoring the measurement model and assuming measurement invariance without verifying whether or not this assumption holds, the accuracy of growth estimates can also be hampered by the omission of respondents missing data. I will turn to this issue in the following section.

C. Respondents with missing data are left out of the analysis

Longitudinal data invariably contain missing data in that not all respondents participate in every wave. Some students cease their studies, while some of their persisting peers miss out on one or more of the data collections (Vermunt et al., 2014). This amount of missing data is a major issue: for the five longitudinal studies the percentage of missing data ranges from 43% (Van der Veken et al., 2009) to 92.2% (Busato et al., 1998).

The current practice is to retain only those respondents with complete data for the analysis of change over time (henceforth named listwise deletion, e.g., Busato et al., 1998; Donche et al., 2010; Van der Veken et al., 2009). However, the methodological literature has repeatedly found that listwise deletion leads to inaccurate estimates (Enders, 2001; Enders & Bandalos, 2001; Wothke, 2000).

First, leaving a large part of the sample out of the analysis (e.g., students for whom information is available on two out of three waves), decreases the sample size. As noted above, in longitudinal studies on the change in learning strategies, the amount of missing data is large, bringing with it a substantial decrease in sample size. This sample size is an important element for statistical power; if fewer cases are included in the analysis, the power to detect effects decreases (Allison, 2009; Cheema, 2014; Peugh & Enders, 2004). Thus, low statistical power can hamper the validity of the conclusions in that results may point towards the absence of growth over time, whereas in fact it existed.

Second and more harmful however, is that the group of students with complete data is very unlikely to be a random subset of those starting higher education (Raudenbush, 2001a). The findings on the relationship between learning strategies and drop-out (e.g., Vanthournout et al., 2012; Watkins & Hattie, 1985) and between learning strategies and non-response (Watkins & Hattie, 1985), stress that students in a longitudinal study of the growth in learning strategies are not missing in a random fashion (Richardson, 2013): they differ in terms of their learning strategies. The sample of students with complete data is thus selective. Conducting an analysis on this sample can produce inaccurate estimates of the growth trend in learning strategies (Wothke, 2000). Consequently, the validity of the conclusions can be compromised (Foster, Fang, & Conduct Problems Prevention Research Group, 2004; Shadish et al., 2002). The results may indicate an absence of growth, whereas, in fact there was, the strength of growth can be assessed incorrectly, or growth over time may be estimated, whereas, in fact there was no growth.

In the light of statistical power and, to account for a selective follow-up sample, the methodological literature recommends other, more modern techniques to deal with missing data, characterised by including respondents with missing data in the analysis (Allison, 2009; Enders, 2010). A number of modern missing-data techniques assuming missing at random (MAR, the assumption that missingness is related to one of the variables in the dataset such as the score on a learning strategy at the first wave) have been described in the literature. Next to this, recently, techniques assuming missing not at random (MNAR, the assumption that missingness is related to the unobserved change over time) have been introduced in the social sciences (Enders, 2011; Muthén & Muthén, 2010). To the best of our knowledge, in the educational sciences, studies comparing the results obtained from techniques assuming MAR and MNAR next to the commonly used listwise deletion, are absent. In the literature on learning strategies for certain, the question regarding whether the missing data technique matters in the light of the results on the growth trend is, up to the present, unanswered.

1.4. Focus and overview of this dissertation

The main goal of the current work is *to model the change in learning strategies during and after the transition to higher education and, to examine the impact of statistical choices on growth estimates.* This consists of two parts. From a substantive focus, the aim is to map the growth trend in learning strategies during the transition from secondary to higher education and during higher education. Here, I discern two research questions:

1. How do students' learning strategies change on average?
2. Is there differential change in learning strategies?

From a methodological focus, I explore the degree to which statistical choices affect the growth trend estimates. Here, three research questions are formulated:

3. Do the estimates of the growth trend in learning strategies differ when the measurement model is taken into account?
4. Can measurement variance over time affect the growth trend estimates?
5. Do growth estimates differ according to the missing data technique adopted?

We note that research questions three to five are relevant for all researchers investigating longitudinal change with self-report Likert-type questionnaires. Regardless of the research topic, researchers setting out to estimate longitudinal change need to make statistical choices on how to handle measurement error, whether to test for LMI, as well as how to deal with missing data.

To answer the five research questions noted above, two samples were studied. The first sample consisted of students during the transition period from secondary to higher education. In this sample, as shown in Figure 1.1., five data waves are included. In the last year of secondary education, students filled out the Inventory of Learning Styles – Short Version (ILS-SV, Donche & Van Petegem, 2008) on two occasions. In the first year of higher education, students were asked to participate anew on two occasions. The last wave took place in December of the second year of higher education.

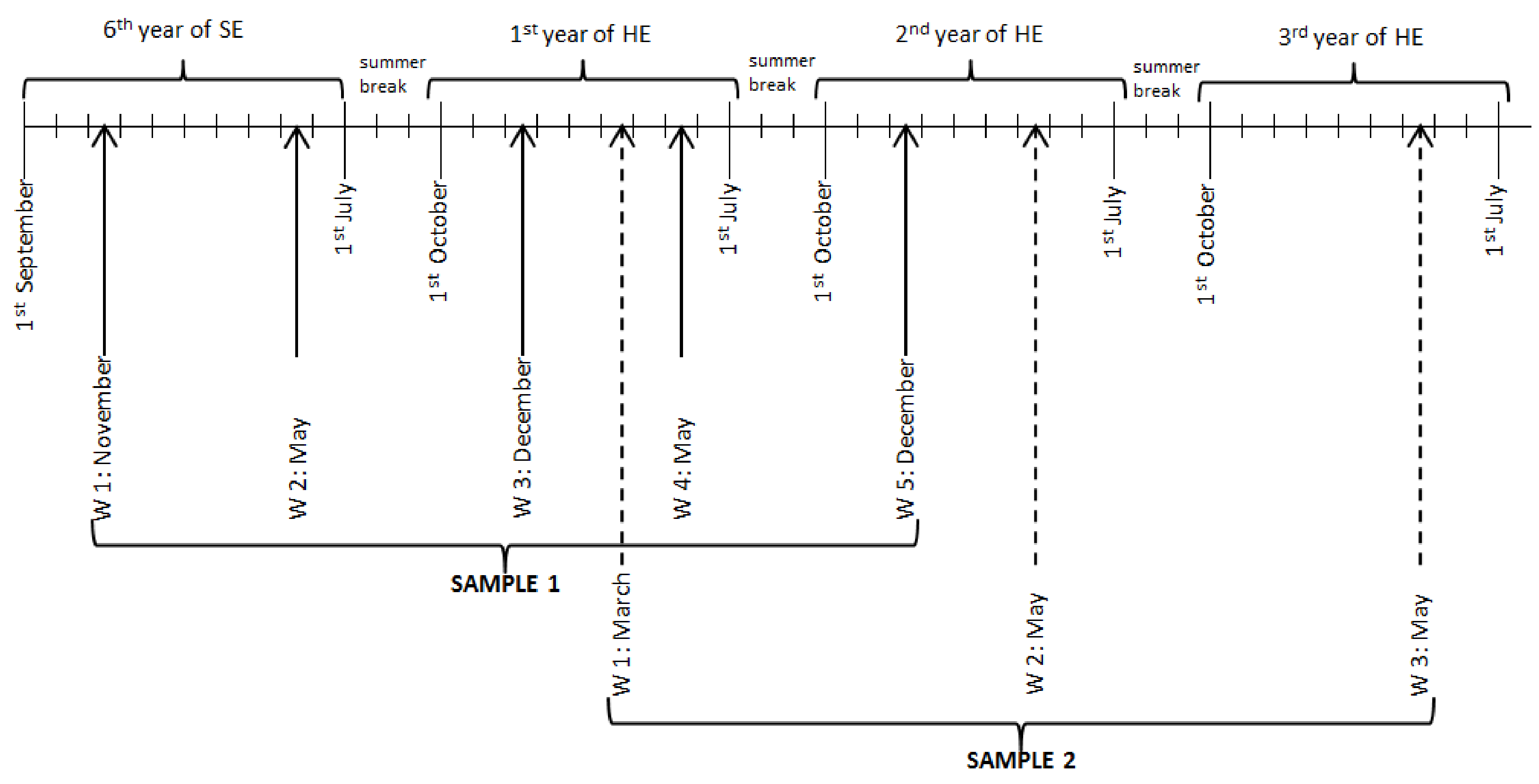


Figure 1.1: Overview of the measurement waves for the two samples

(SE=Secondary Education, HE=Higher Education, W=wave)

For the second sample, one cohort of students in a Flemish University College was followed over three waves (see Figure 1.1). In March of the first academic year, all first-year students were asked to fill out the ILS-SV during scheduled lecture slots. The same cohort was questioned again in May of the second and the third year.

These two samples have been studied over the course in five chapters of the dissertation. Table 1.3 provides an overview of these chapters, the research questions on which they focus, as well as the sample relied upon. Prior to detailing these chapters, I note that Chapters Two to Four have been published in an international journal. Chapters five and six are currently under review. A benefit of a dissertation based on articles is that each chapter can be read separately. A downside of this structure is that chapters overlap to some extent, such as, for example in describing the sample or the measurement instrument. In the next paragraphs, I will shortly explore the focus of each of the chapters.

Chapter 2: Taking the measurement model into account

Up to the present, repeated measures ANOVA has been mostly relied upon to map longitudinal change in learning strategies. Another technique, multilevel modelling, has been recently used as well (e.g., Vanthournout, 2011). However, neither analytic technique allows the researcher to take the measurement model into account. On the other hand, a MILG model does. In chapter two, by using the three analysis techniques on the data of sample 1, we investigate whether and how taking the measurement model into account affects the estimates of average and differential growth.

Chapter 3: Longitudinal measurement invariance testing

An additional advantage of MILG is that it allows the researcher to verify whether or not the learning strategy questionnaire measures equivalently over the different data waves. Testing this is crucial to disentangling true change from change in the measurement over time. The third chapter showcases how measurement invariance can be tested for, using the data of sample 2. Next to this, depending on the severity of the violation of the measurement invariance hypothesis, different recommendations for the subsequent modelling of growth over time are formulated.

Chapter 4: The growth trend in learning strategies during higher education

In the fourth chapter, the data of sample 2 are analysed from a substantive focus: the average and differential growth in learning strategies during higher education are described. In doing so, recommendations from chapter two are followed: the measurement model underlying the scale scores is acknowledged. The findings from chapter three on the measurement variance over time are taken into account as well (see Table 1.3).

.

Table 1.3: Overview of chapters, research questions (RQ) and samples

	Ch. 2	Ch. 3	Ch. 4	Ch. 5	Ch. 6
RQ1: How do students' learning strategies change on average?			*Focus* (HE)	*Focus* (transition to HE)	
RQ2: Is there differential growth in learning strategies?			*Focus* (HE)	*Focus* (transition to HE)	
RQ3: Do the estimates of the growth trend differ when the measurement model is taken into account?	Focus		Applied	Applied	
RQ4: Can measurement variance over time affect the growth trend estimates?		Focus	Applied	Applied	
RQ5: Do growth estimates differ according to the missing data technique adopted?				(Partially) Applied	Focus
Sample 1	X°			X	
Sample 2		X	X		X

° the first three out of five waves were used in this study

Chapter 5: The growth trend in learning strategies during the transition from secondary to higher education

In addition to examining the change in learning strategies in a stable educational context as done in chapter four, we set out to investigate the growth of learning strategies when there is a transition in the educational context. The fifth chapter focuses on mapping the average as well as differential growth in learning strategies in the transition period from secondary to higher education. In doing so, recommendations regarding the measurement model, measurement invariance testing, and missing data, are followed

Chapter 6: Sensitivity of the growth trend estimates to the missing data technique

In chapter five, respondents with partially missing data are retained for the analysis. The sixth chapter takes this one step further: the range of modern missing-data techniques is showcased, and sensitivity analysis is conducted on the growth trend during higher education (sample 2). With this, the sixth chapter investigates whether or not the missing data technique adopted influences the growth trend estimates on the one hand, and the substantive conclusions drawn from these results on the other.

Chapter 7: Conclusion and discussion

In the seventh and final chapter, the five research questions of this dissertation are answered supported with evidence from chapters two to six. Subsequently, the limitations of the current work and general directions for further research are presented. To conclude, the implications of this research for research practice on the one hand, and for policy and practice on the other, are provided.

2. Taking the measurement model into account

This chapter is based on Coertjens, L., van Daal, T., Donche, V., De Maeyer, S., Vanthournout, G., & Van Petegem, P. (2013). Analysing change in learning strategies over time: A comparison of three statistical techniques. *Studies in Educational Evaluation, 39*, 49-55. doi: 10.1016/j.stueduc.2012.10.006

Liesje Coertjens and Tine van Daal contributed equally to this article as first authors.

Change in learning strategies during higher education is an important topic of research when considering students' approaches to learning. Regarding the statistical techniques used to analyse this change, repeated measures ANOVA is mostly relied upon. Recently, multilevel and multi-indicator latent growth (MILG) analyses have been used as well. The present study provides details concerning the differences between these three techniques. By applying them to the same dataset, we aim to answer two research questions. Firstly, how are findings on the average trend complementary, convergent or divergent? Secondly, how are results on the differential growth over time complementary, convergent or divergent? Data originates from a longitudinal study on the change in learning strategies during the transition from secondary to higher education in Flanders (Belgium). 425 students provided complete data at each of the three waves of data collection. Results on the significance of average trends are convergent while the strength of the growth over time diverges across analysis techniques. Regarding the differential change, the MILG seems more able to detect variance in growth over time. Recommendations for future research on the changeability of learning strategies over time are provided.

2.1 Introduction

Since the end of the nineteen eighties, student learning has been intensively studied, in particular in higher education contexts. This resulted in a vast array of style models mapping individual differences in student learning (Coffield, Mosley, Hall, & Ecclestone, 2004). Of those, the students' approaches to learning (SAL) model maps students' deep and surface processing (Biggs et al., 2001; Entwistle, McCune, & Hounsell, 2003). The learning pattern model on the other hand distinguishes four patterns: meaning-oriented, reproduction-oriented, application directed and undirected (Vermunt & Vermetten, 2004). Both models start from the viewpoint that there are individual differences in going about learning which are associated with learning outcomes. Moreover, both models take into account that different ways of processing learning contents also goes along with differences in study motivation and rely upon students' perceptions of their learning (Desmedt & Valcke, 2004; Lonka et al., 2004; Vanthournout, 2011). Therefore, the concept SAL has recently been used to encompass both models (Vanthournout, 2011).

From the research interest of the consistency or variability in learning approaches, longitudinal studies have increased in recent years (Phan, 2008; Vanthournout et al., 2011). Using, for example, the Study Process Questionnaire (SPQ; Biggs et al., 2001), Approaches to Study Inventory (ASI; Entwistle & Ramsden, 1983) or Inventory of Learning Styles (ILS; Vermunt & Vermetten, 2004), researchers have assessed changes in learning approaches by using three or more measurements of a learning approach questionnaire (e.g., Reid et al., 2005; Van der Veken et al., 2009; Zeegers, 2001).

Examining the findings of these longitudinal studies, relying on Biggs and colleagues' framework and the SPQ, Gordon and Debus (2002) and Jackling (2005) noted an increase in deep processing over time, while Zeegers (2001) concluded it to remain constant. Concerning surface processing, Gordon and Debus (2002) detected a decreasing reliance, while the other two studies found a constant trend. Using the Approaches to Study Skills Inventory for Students (ASSIST, Tait, Entwistle, & McCune, 1998), an adaptation of the ASI, Reid and colleagues (2005) noted a decrease in both the deep and the strategic approach over the first year of medical training, while the surface approach did not alter. Over the second year, the surface and deep approach remained stable, while the strategic approach continued to decrease.

Five longitudinal studies use the Vermunt framework and (parts of the) ILS to map changes throughout higher education (Busato et al., 1998; Donche et al., 2010; Severiens et al., 2001; Van der Veken et al., 2009; Vanthournout, 2011). Concerning meaning-oriented learning (Busato et al., 1998) or deep processing strategies (Donche et al., 2010; Severiens et al., 2001; Van der Veken et al., 2009) an increase was found. However, Vanthournouts study (2011) suggests that only the subscale relating and structuring scale increases, while the other subscale of deep processing, critical processing, remained constant. Stepwise processing or its subscale analysing was found to remain constant over time (Busato et al., 1998; Donche et al., 2010; Severiens et al., 2001; Vanthournout, 2011), while the degree of memorizing decreased over time (Donche et al., 2010) or showed a quadratic trend with a rise after an initial decrease (Vanthournout, 2011). Concerning the regulation strategies, self-regulation was mostly found to increase (Donche et al., 2010; Severiens et al., 2001; Vanthournout, 2011), while remaining constant in Van der Veken et al.'s (2009) study. External regulation on the other hand decreased (Donche et al., 2010; Severiens et al., 2001; Vanthournout, 2011). Lastly, undirected learning was found to remain constant (Busato et al., 1998; Severiens et al., 2001; Van der Veken et al., 2009) or to decrease over time (Donche et al., 2010; Vanthournout, 2011).

Examination of the statistical analysis of these longitudinal studies within the students' approaches to learning (SAL) field reveals a strong reliance on repeated measures ANOVA. Recently, two other techniques have demonstrated their value when conducting research in this domain: multilevel modelling and multi-indicator latent growth (MILG) analysis. In this study, we aim to understand how these three techniques can be valuable to address the changeability of learning strategies over time. We explore this threefold perspective on the same database and discuss the differences in data treatment and subsequent findings.

Regardless of the research domain tackled here, the findings of this study may also be of interest to researchers in other educational research fields. When change over time is investigated, for example to assess the impact of an educational program on students cognitive or metacognitive abilities (e.g., Breuer & Eugster, 2006), insight into techniques to properly estimate such change is important. In what follows, we first describe the goals, analysis procedure and assumptions of the three techniques. Next, we discuss the research findings of an illustrative empirical study, demonstrating to which degree the analytical techniques provide complementary, convergent or divergent information on the nature of change in learning strategies.

2.1.1 Repeated measures ANOVA

Repeated measures ANOVA is the most frequently reported technique to examine growth in learning strategies over time (Donche et al., 2010; Nienemin, Lindblom-Ylänne, & Lonka, 2004; Severiens et al., 2001; Smith et al., 2007; Zeegers, 2001). The goal of this technique is to discern whether scale scores vary over time. The null hypothesis states that the scale scores are equal over time. When the null hypothesis is rejected this indicates that at least two scores differ.

To conduct a repeated measures ANOVA, three steps are usually taken. First, manifest scale scores are computed. Per student, the scores on the items for each scale are averaged at each wave of data collection. Subsequently, repeated measures ANOVA are relied upon to compare the mean manifest scale scores over time. Third, when a significant difference in manifest scale scores over time is detected, post hoc tests are applied. On the condition that data collection is spaced equally over time, trend analysis relying on polynomial contrasts can be used (Green & Salkind, 2003), testing to which degree a certain trend (e.g. linear or quadratic) fits the observed pattern in manifest scale scores over time.

Repeated measures ANOVA hinges upon a number of assumptions (Green & Salkind, 2003), two of which are relevant in comparison to multilevel and MILG analysis. Firstly, repeated measures ANOVA is based upon the sphericity assumption stating that variances of the difference scores are presumed equal over time. Statistical programs however offer multivariate tests which are robust to violations of this assumption (Green & Salkind, 2003). Secondly, it is assumed that manifest scale scores can be relied upon to draw conclusions with regard to latent factors (e.g., deep learning decreases over time). The measurement error associated with each item is not taken into account.

2.1.2 Multilevel modelling

Recently, a number of studies on the changeability in learning strategies over time has applied multilevel analysis (e.g., Endedijk, 2010; Vanthournout, 2011).

In such analysis, it is assessed whether a growth trajectory (e.g., linear, quadratic, cubic, etc.) holds (Heck, Scott, & Tabata, 2010; Singer & Willet, 2003). Taking the simplest case, a dataset with three measurement waves, a linear growth model can be estimated. Multilevel analysis then estimates an intercept (the average score for a scale at the first wave) and a slope (the average increase or decrease in the scale scores per one unit of time, e.g. one year) and indicates whether these parameters differ significantly from zero. If the null hypothesis is rejected for the slope variable this implies that a linear growth trajectory over time fits the data significantly better than a constant growth trajectory.

Next to assessing whether a certain growth trajectory holds, multilevel analysis also aims at modelling individual variations between students. By taking into account that observations are nested within the respondents, three additional parameters provide information on this differential growth (Hox, 2000; Singer & Willet, 2003; Voelkle, Wittmann, & Ackerman, 2006). First, the intercept variance parameter expresses whether students vary significantly in their initial level on a learning strategy scale. Second, the variance in slopes indicates whether students follow the general trend or whether they deviate from one another. Third, the covariance parameter indicates whether students' initial scores on a learning strategy scale are related to their change over time.

Multilevel analysis consists of three steps (Heck et al., 2010). First, manifest scale scores are computed for each scale. Second, a random intercepts (RI) model is estimated, assessing whether for example a linear growth trajectory holds and whether students differ in their initial value on a learning strategy scale (intercept variance). Third, if the slope parameter is significant, a random slopes

(RS) model is relied upon to assess whether students vary in their growth over time. The slope variance and covariance parameters are then estimated.

A number of assumptions underpin multilevel analysis (Hox, 1998). Regarding the comparison of the three techniques, and comparable to repeated measures ANOVA, we note one: manifest scale scores are relied upon, implying that measurement error associated with each item is not modelled explicitly. A MILG analysis does take this error into account.

2.1.3 Multi-indicator latent growth (MILG) analysis

A less often applied technique in assessing the changeability in learning strategies over time is MILG analysis (Coertjens, Donche, De Maeyer, Vanthournout, & Van Petegem, 2013a). The goals of this technique are identical to multilevel analysis: average as well as differential growth is estimated (Byrne, 2010). The former is assessed by testing whether a growth trajectory holds. Concerning the latter, the intercept variance, slope variance and covariance parameters are estimated.

Contrary to multilevel analysis, measurement error is taken into account when assessing growth (Byrne, 2010). A latent scale score is estimated per measurement wave and on these scale scores the parameters of the average and differential growth trajectory are estimated (Metha, Neale, & Flay, 2004; Wu et al., 2010). The MILG analysis estimates the elements of the average (e.g., intercept and slope) and of the differential growth trajectory (intercept variance, slope variance and covariance) in a single step.

2.1.4 The current research

To our knowledge, no study in the SAL domain has critically examined the value of repeated measures ANOVA, multilevel analysis and MILG analysis in assessing the changeability of learning strategies over time. Given the various goals, analytical procedures and assumptions, the question arises to whether these three techniques suggest complementary, convergent or divergent findings.

This study therefore applies the three analysis techniques to the same non-simulated dataset. Two research questions are put forward. Firstly, how are findings on the average trend complementary, convergent or divergent? To answer this research question, the results from the repeated measures ANOVA, multilevel and MILG analysis are compared. Secondly, how are the results on the differential growth over time complementary, convergent or divergent? To answer the second research question, the results from the multilevel models and MILG models are contrasted.

Over the last few decades, a lot of research effort has been invested in exploring the ways in which students learn in higher education. This research stems from a variety of research traditions (Lonka et al., 2004; Richardson, 2007) and has evolved in different directions. A large number of studies have been carried out in diverse areas, such as: cognitive aspects of learning (Moskvina & Kozhevnikov, 2011); learning styles (Kolb, 1984); intellectual styles (Zhang & Sternberg, 2005); learning conceptions (Van Rossum & Schenk, 1984), approaches to learning (Marton & Säljö, 1997); aspects of self-regulation (Boekaerts, Pintrich,

& Zeidner, 2000); study orientations (Nieminen, Lindblom-Ylänne, & Lonka, 2004; Richardson, 1997); meta-cognition (Flavell, 1987); and motivational aspects of learning (Boekaerts & Martens, 2006). A shared feature of many of these studies is the search for relationships between various aspects of learning and an attempt to arrive at integrative models of learning (Biggs, 1993; Entwistle & McCune, 2004; Meyer, 1998; Vermunt & Vermetten, 2004).

One of the research traditions interested in student learning in higher education is the Students' Approaches to Learning-tradition (the SAL tradition; Lonka et al., 2004). It is founded on the phenomenographical studies by Marton and Säljö in the seventies of the previous century (Marton & Säljö, 1976). Research in this tradition generally focuses on the different ways students engage in learning or handle learning tasks as reported by the students themselves (Biggs, 2001; Entwistle, McCune, & Scheja, 2006; Schmeck, 1988). Representatives of this tradition mostly concur on the viewpoint that there are qualitatively different ways in which students go about learning and that these differences in learning approaches are associated with qualitatively different learning outcomes (Biggs, 1979; Entwistle, Meyer, & Tait, 1991; Richardson, 1997; Vermunt, 2005). How students approach their learning is viewed as being influenced by factors in the learning environment, students' perceptions of these factors and student characteristics (Baeten, Kyndt, Struyven, & Dochy, 2010; Biggs, 2003; Donche & Van Petegem, 2006; Vermunt, 2005).

2.2 Method

2.2.1 Design and respondents

The data stem from a research project on students' transition from secondary to higher education. A group of 5 technical and vocational education Flemish schools interested in their performance requested the research. To be able to provide an adequate picture on this performance, 11 schools were matched on their study domains, thus generating a purposive sample of 16 schools offering technical and vocational education. Next to this, a weighted random sample of 20 general education schools was drawn. In the 36 schools, all final year students were selected for the study.

During the last year of secondary education, students were questioned twice during school hours, in November and May. At the second wave, students were asked to provide us with their contact information (email, home address and telephone number). In December of the second year, students were invited to participate anew. In total 425 students were studying in higher education and provided complete data at each of the three waves. Of those 425, 299 students came from general education while 126 students had a degree from technical or vocational education. Due to the shift in educational context that the students underwent, a change in learning strategies is hypothesized. The data consist of three waves, being the simplest possibility for which the three techniques can be demonstrated and being common in longitudinal research on the change in learning strategies (Vanthournout et al., 2011).

2.2.2 Measurement

Learning strategies are investigated by focusing on two malleable components of the learning patterns of the Vermunt framework: cognitive processing and regulation strategies (Vermunt, 1998). The scales used in this study stem from the 'Inventory of Learning Styles – Short Version' (ILS-SV), which has been validated for first-year Flemish University College students (Donche & Van Petegem, 2008). Processing strategies can be viewed as the cognitive activities a student applies whilst studying. In the ILS-SV, four scales for cognitive processing strategies are distinguished: memorizing, analysing, critical processing and relating and structuring. Regulation strategies are metacognitive activities that students undertake, such as planning or testing oneself. To map regulation strategies, the ILS-SV discerns three scales: external regulation, self-regulation and lack of regulation. For all seven scales, the items are scored ranging from (1) 'I never or hardly ever do this' to (5) 'I (almost) always do this'. Table 2.1 provides for each scale the number of items and an example item.

Since Cronbach alpha values are very sensitive to the number of items (Cortina, 1993; Schmitt, 1996) and the ILS-SV scales consist of only 4 to 6 items, reliability analysis was conducted using the mean inter-item correlation. To discern good reliability a cut-off of .20 is put forward (Palant, 2007). For the external regulation scale at the first wave the value was close to the cut-off (.18, see Table 2.1), while exceeding it for the second and third wave. For all other scales at each wave the criterion for good reliability was met.

Table 2.1: Learning strategy scales of the ILS-SV questionnaire, number of items, item examples (translated from Dutch) and mean inter-item correlation

Scales	Nr. of items	Item example	Mean inter-item correlation
Processing strategies			
Memorizing	4	I learn definitions by heart and as literally as possible.	.33-.40
Analysing	4	I study each course book chapter point by point and look into each piece separately.	.28-.32
Critical processing	4	I try to understand the interpretations of experts in a critical way.	.38-.43
Relating and structuring	4	I compare conclusions from different teaching modules with each other.	.33-.37
Regulation strategies			
External regulation	6	I study according to the instructions given in the course material.	.18-.24
Self-regulation	4	I use other sources to complement study materials.	.31-.35
Lack of regulation	4	I confirm that I find it difficult to establish whether or not I have sufficiently mastered the course material.	.37-.42

2.2.3 *Plan of analysis*

Table 2.2 reports the mean manifest scale scores and standard deviations at each wave. These data are analysed using three statistical methods. First, repeated measures ANOVA are estimated in SPSS relying upon the mean manifest scale scores per wave. To assess the size of the effect, the partial eta square is reported, for which the values .01, .06 and .14 are interpreted as being respectively small, medium and large (Green & Salkind, 2003). Subsequently, given that the waves are almost equally spaced (6 months between wave 1 and 2 and 7 months between wave 2 and 3), trend analysis is conducted (Green & Salkind, 2003). This post hoc test assesses whether the found change in manifest scale scores over time best fits a linear or quadratic trend.

Table 2.2: Mean manifest scale scores and standard deviation for the learning strategy scales

	1st wave	2nd wave	3rd wave
Memorizing	3.445 (0.809)	3.374 (0.809)	3.499 (0.782)
Analysing	3.258 (0.740)	3.242 (0.717)	3.470 (0.678)
Critical processing	2.982 (0.814)	2.975 (0.825)	3.288 (0.752)
Relating and structuring	3.088 (0.712)	3.177 (0.704)	3.529 (0.656)
External regulation	3.570 (0.576)	3.482 (0.623)	3.715 (0.560)
Self-regulation	2.205 (0.732)	2.242 (0.767)	2.758 (0.761)
Lack of regulation	2.191 (0.764)	2.236 (0.818)	2.668 (0.768)

Second, transforming the dataset into a person-period file allows multilevel analysis in SPSS (Singer & Willet, 2003). The three data points are labelled 0, 0.5 and 1.08 respectively so that the slope parameter provides the change in a learning strategy over one year. The RI model, estimates the intercept (i.e., average initial value for the scale, in our case the value in October of the last year of secondary education), slope (i.e., average increase in the scale per one increase of time, here 1 year) and intercept variance (indicating whether students differ in their initial values on the scale). If the slope parameter results significant, a RS model is estimated to assess whether students vary in their growth trajectory. Results of this RS model are reported when the RS model fits the data significantly better than the RI model (in terms of the Chi^2 difference test) and when either the slope variance or the covariance reaches the significance cut-off point.

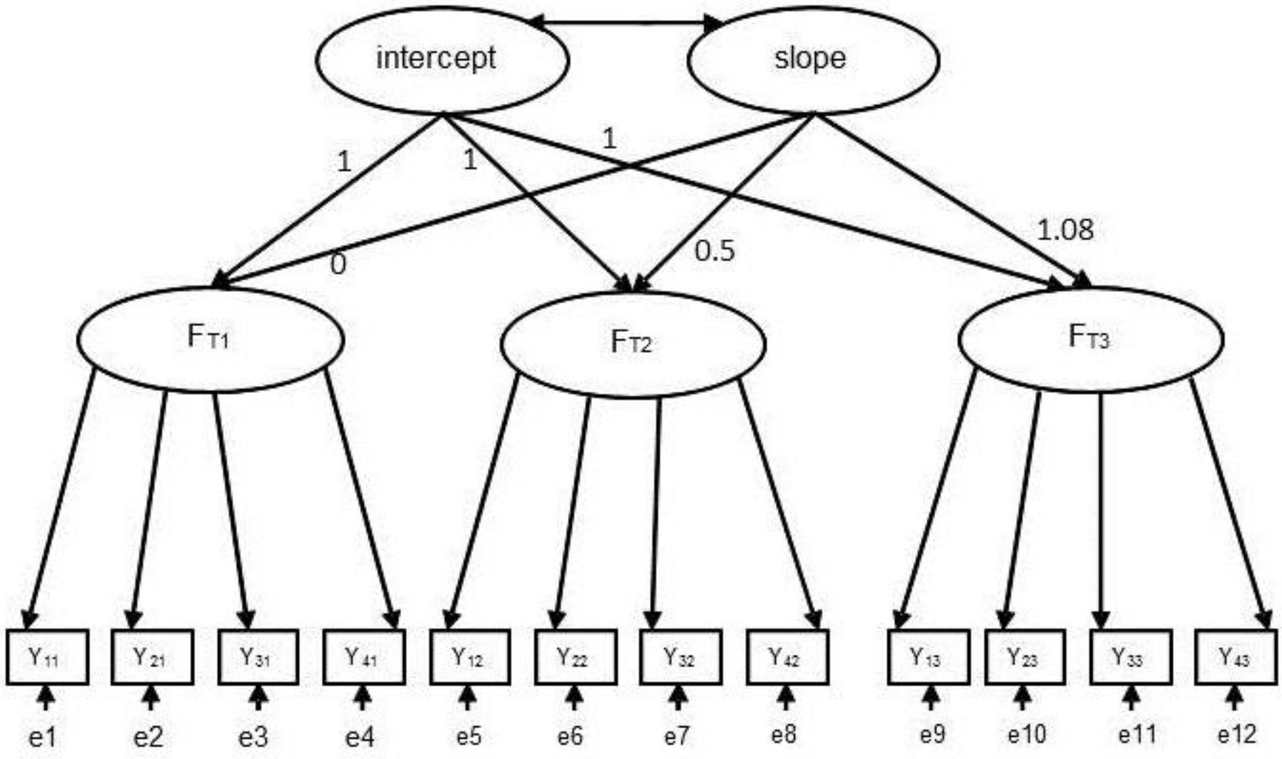


Figure 2.1: MILG model

A third analysis is the MILG model (Muthén & Muthén, 2010). Figure 2.1 exemplifies such a model, consisting of two levels. A factor (e.g., the latent concept memorizing) is measured at three moments, using the same four items each time (Y_1-Y_4)[1]. An individual's score on an item at a certain time (Y_{ijt} e.g. Y_{i31} represents the score for individual i on the third item at the first wave) is predicted by a latent factor (e.g. F_{T1}). The second-order factors - intercept and slope - serve to explain the mean and covariance structure of these latent factors (Hox, 2000; Stoel, Roeleveld, Peetsma, van den Wittenboer, & Hox, 2006). The values of the factor loadings for the slope are adjusted to 0, 0.5 and 1.08

[1] Figure 2.1 depicts the situation for all scales except the external regulation scale, having not 4 but 6 items (see Table 2.1).

respectively. The interpretation of the intercept, slope and three variance parameters is comparable to multilevel analysis. MILG analyses were performed in Mplus 6.1 using a maximum likelihood estimation procedure.

2.3 Results

Table 2.3 gives the results of the repeated measures ANOVA, partial eta² and post hoc tests. Tables 2.4 and 2.5 depict the results of the multilevel and MILG analysis respectively. First, we will examine the results for the average trend per learning strategy scale. Second, we will detail the differential growth estimates of the multilevel and MILG model.

2.3.1 Average growth

For all scales, the repeated measures ANOVA analysis suggests that the manifest scale scores cannot be considered equal across the three waves (see Table 2.3). The size of the effects is judged small for the memorizing scale, medium for the analysing scale, while large for the critical processing, relating and structuring, external, self- and lack of regulation scales.

Examining the results of the post hoc tests reveals a significant linear estimate for six out of the seven scales and a significant quadratic estimate for all scales (see Table 2.3). Concerning the critical processing, relating and structuring, self-regulation and lack of regulation scales, the effect sizes suggest the linear trend is more powerful in explaining the change in the manifest scales scores over time. Regarding the analysing and external regulation scales, neither the effect size for the linear nor for the quadratic time effect is large. For the memorizing scale, a quadratic trend fits the shape in the manifest scale scores, but the effect size suggests the degree of explained variance is small.

Multilevel and MILG analysis detect a significant linear increase for six of the seven learning strategy scales (see Table 2.4 & 2.5). For each of these scales, the slope is significant and positive. For the memorizing scale on the other hand, results from the analysis techniques do not confirm a linear effect: the slope is not significantly different from zero.

Though the multilevel and MILG analysis suggest comparable conclusions concerning whether a linear effect holds, the magnitude of the change over time varies significantly for two scales (see Table 2.4 & 2.5). For the external regulation scale, the 95% confidence interval of the slope in the multilevel model has a lower and upper limit of respectively 0.091 and 0.193, while the interval spans from 0.220 to 0.369 for the MILG model. Using MILG analysis, the slope is thus estimated significantly steeper. The confidence intervals of the slopes for the lack of regulation scale do not overlap either (0.385 to 0.515 and 0.259 to 0.369 for the multilevel and MILG model respectively), suggesting the MILG analysis estimates the slope significantly lower.

Table 2.3: Results of one-way repeated-measures ANOVA and trend analysis

	Wilks′ Δ	Partial eta²	Linear		Quadratic	
			p-Value	Partial eta²	p-Value	Partial eta²
Memorizing	.970 **	.030	ns	.006	**	.024
Analysing	.882 ***	.118	***	.079	***	.042
Critical processing	.814 ***	.186	***	.143	***	.064
Relating and structuring	.701 ***	.299	***	.279	***	.054
External regulation	.852 ***	.148	***	.063	***	.100
Self-regulation	.577 ***	.423	***	.378	***	.143
Lack of regulation	.665 ***	.335	***	.306	***	.093

*** p<.001; ** p<.01; * p<.05

Table 2.4: Parameter estimates for the multilevel models

	Intercept	Slope	Intercept variance	Slope variance	Covariance
Memorizing	3.410 (.037)***	.054 (.033)	.376 (.032) ***		
Analysing	3.217 (.033)***	.201 (.032) ***	.260 (.024) ***		
Critical processing	2.929 (.037)***	.289 (.033) ***	.368 (.032) ***		
Relating and structuring	3.047 (.034)***	.413 (.032) ***	.209 (.014) ***	.074 (.039)	-.083 (.029) *
External regulation	3.514 (.027)***	.142 (.026) ***	.185 (.017) ***		
Self-regulation	2.137 (.035)***	.523 (.032) ***	.321 (.028) ***		
Lack of regulation	2.218 (.037)***	.450 (.033) ***	.358 (.031) ***		

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

Table 2.5: Parameter estimates for the MILG models

	Slope	Intercept variance	Slope variance	Covariance
Memorizing	.026 (.029)	.357 (.053) ***	.084 (.065)	-.051 (.038)
Analysing	.215 (.036) ***	.329 (.053) ***	.083 (.074)	-.080 (.046)
Critical processing	.282 (.036) ***	.498 (.068) ***	.167 (.082) *	-.146 (.053) **
Relating and structuring	.363 (.032) ***	.309 (.044) ***	.139 (.054) *	-.121 (.036) **
External regulation	.295 (.038) ***	.471 (.066) ***	.256 (.086) **	-.178 (.055) **
Self-regulation	.531 (.040) ***	.372 (.057) ***	.053 (.094)	-.030 (.054)
Lack of regulation	.314 (.028) ***	.232 (.034) ***	.071 (.043) *	-.055 (.026) *

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

2.3.2 *Differential growth*

Looking at the differential trends as estimated by the multilevel and MILG models, both analysis techniques indicate significant intercept variance for all scales. Only for the lack of regulation and memorizing scale this variance estimate is smaller using the MILG analysis. Relying on 95% confidence intervals of the variance estimates, the difference is judged significant for four scales. Compared to the multilevel model, the MILG analysis estimates the intercept variance to be larger for the critical processing, relating and structuring and external regulation scales, while smaller for the lack of regulation scale.

There are also differences for the slope variances and the covariance. Regarding the six scales for which these indicators were relevant (excluding the Memorizing scale which did not show significant change over time), two scales showed similar estimates. Concerning the self-regulation and analysing scale, both analysis techniques indicate that the general trend for this scale can be assumed valid for all students (see Table 2.4 and 2.5).

For three scales (critical processing, external regulation and lack of regulation), the multilevel analysis indicated that students follow a comparable growth trajectory over time (see Table 2.4), while the results from the MILG analysis detected significant slope variance (see Table 2.5). For these scales, the MILG model suggests that students vary in their change over time (significant slope parameters) and that students scoring lower at the first wave increase their reliance on these strategies more rapidly than students scoring higher (significant negative covariance).

Lastly, for the relating and structuring scale, the multilevel model indicated a significant covariance estimate, while the slope variance did not reach the significance cut-off level. The estimates from the MILG model on the other hand, point towards a significant slope variance and covariance, suggesting anew that students scoring higher at the first wave increase their relating and structuring to a lesser degree.

2.4 Conclusion and discussion

Longitudinal research studies on the change in students' approaches to learning are on an increase. From a statistical point of view, a repeated measures ANOVA is mostly relied upon to assess change in learning strategies over time in the SAL field. Compared with this technique, multilevel and MILG analyses are scarcely applied. The present study therefore illustrated the goals, analytical procedure and assumptions underpinning these techniques. Moreover, by applying the three techniques to the same longitudinal dataset, we examined whether and how the three techniques result in complementary, convergent or divergent findings on the average and differential change in learning strategies.

Results on the average change indicate that the three techniques are convergent on the significance of linear growth trajectories. However, for some scales the multilevel and MILG models diverge concerning the strength of this change over time. Regarding the differential change, results converge anew on the significance of the intercept variance, while diverge on their magnitude. The results of the multilevel model and the MILG model are however divergent on the slope variance and the covariance.

The substantive conclusions on the average growth trajectories are comparable across analysis techniques for six of the seven processing and regulation strategy scales. Analysing, critical processing, relating and structuring, external regulation, self-regulation and lack of regulation were all found to increase over time. For the memorizing scale however, repeated measures ANOVA suggests that the manifest scale scores differ over time and that this change over time is best - though not strongly - explained by a quadratic trend. Results from the multilevel and MILG techniques indicate that a linear trend does not fit the data. This illustrates that the last finding should not automatically be interpreted as absence of change (i.e., a constant trend) over time. An alternative explanation is that the change over time follows a more complex growth trajectory in reality. With three data points, multilevel and MILG analysis can only assess constant or linear growth trajectories (Metha et al., 2004; Wu et al., 2010). In sum, the reported results on the average growth are only seemingly divergent. To capture whether a constant trend is likely to reflect an absence of change, inspection of the mean manifest scale scores (see Table 2.2) prior to multilevel analysis or MILG can be put forward as a good practice.

Though speaking in one voice concerning the significance of linear growth trajectories, the multilevel and MILG models differ substantially concerning the strength of this change over time for two scales in this illustrative empirical study. This result is at odds with prior findings by Hox (2000) and Curran (2003). For the external regulation scale, the MILG analysis suggests a steeper slope than the multilevel model. This could be due to the fact that the former takes measurement error into account, which was found to be larger for this scale compared to others (see Table 2.1). For the lack of regulation scale the MILG model estimated the slope to be significantly lower than the multilevel model. The reliability estimates however do not seem to cause this. To our knowledge, the methodological literature does not suggest a possible explanation for this finding. Simulation studies are warranted to clarify the origin of this difference in slopes between the multilevel and MILG model.

The multilevel and MILG models estimate differential growth. Concerning the intercept variance, the latter analysis technique estimates this to be significantly larger for three scales. Anew for the lack of regulation scale, the intercept variance is significantly lower for the MILG analysis compared to the multilevel analysis. The slope variance moreover differs in significance between the techniques. Next to this, the MILG model seems more powerful to detect differential growth over time. For the six learning strategies scales for which slope variance was examined (excluding the memorizing scale due to its constant trend), four scales depicted differential change over time using the MILG model but not in the multilevel model. Simulation studies could be fruitful in clarifying under which conditions (e.g. degree of measurement error, sample size, strength of true effect) MILG models are more powerful in estimating differential growth than multilevel models.

The present study is subject to a number of constraints. First, listwise deletion was used, retaining only those respondents with complete information on each item and at each wave. Contrary to repeated measures ANOVA, multilevel and

MILG models allow for modelling incomplete data (Hox, 2000; Quené & van den Bergh, 2004). To avoid entangling different samples with different analysis techniques, this benefit has not been demonstrated in this study. Given that longitudinal research on the change in learning strategies reports substantial amounts of missing data (e.g., Busato et al., 1998; Donche et al., 2010), it would be worthwhile to explore the inclusion of incomplete data and its effect on growth estimates.

Second, making inferences about change in a learning strategy scale based upon repeated measurement hinges upon the assumption of longitudinal measurement invariance: the definition of the latent construct is required to be comparable over time (Marsh & Grayson, 1994; Metha et al., 2004; Muthén & Muthén, 2009b). Though this assumption holds good for repeated measures ANOVA, multilevel and MILG analysis alike, only the last technique provides the opportunity to falsify this assumption prior to estimating growth over time. Studies with the substantial focus of growth over time thus should preferably include measurement invariance testing prior to estimating growth over time (Coertjens, Donche, De Maeyer, Vanthournout, & Van Petegem, 2012). The next chapter details on how longitudinal measurement invariance can be tested for.

The constraints of the present study notwithstanding, results indicate that the choice of analysis technique has an impact on the resulting evidence concerning the change in learning strategies over time. Given the different goals, analytical procedure and assumptions, repeated measures ANOVA, multilevel and MILG analysis suggested convergent as well as divergent findings on the average and differential growth over time.

Four recommendations for practice can be put forward when assessing change over time, for example when evaluating the longitudinal impact of an educational program. First, examine the mean manifest scale scores over waves prior to assessing change over time. Second, when the data are gathered at unequal time intervals, post hoc tests using trend analysis cannot be relied upon. Multilevel and MILG analysis are however able to model unequal time intervals explicitly (Singer & Willet, 2003). Third, when differential growth over time seems plausible, multilevel and MILG analysis are recommended as well. Of those techniques however, the latter appears more sensitive to detecting such differential change in this study. Fourth, when reliability analysis of the concept under study suggests lower degrees of consistency, (e.g. low Cronbach alpha or mean inter-item correlation), MILG analysis is judged more appropriate.

3. Longitudinal measurement invariance testing

This chapter is based on Coertjens, L., Donche, V., De Maeyer, S., Vanthournout, G., & Van Petegem, P. (2012). Longitudinal measurement invariance of learning strategy scales: Are we using the same ruler at each wave? *Journal of Psychoeducational Assessment*, *30*(6), 577-587. doi: 10.1177/0734282912438844

Whether or not learning strategies change during the course of higher education is an important topic in the Students' Approaches to Learning field. However, there is a dearth of any empirical evaluations in the literature as to whether or not the instruments in this research domain measure equivalently over time. Therefore, this study details the procedure of longitudinal measurement invariance testing of self-report Likert-type scales, using the case of learning strategies. The sample consists of 245 University College students who filled out the Inventory of Learning Styles – Short Version three times. Using the weighted least squares means-variance (WLSMV) estimator to take into account the ordinal nature of the data, a series of models with progressively more stringent constraints were estimated using Mplus 6.1. The results indicate that longitudinal measurement invariance holds for all but two learning strategy scales. The implications for longitudinal analysis using scales with varying degrees of measurement invariance are discussed.

3.1 Introduction

Educational researchers have long been interested in how students learn in higher education. One perspective on this issue is offered by the Students' Approaches to Learning tradition (SAL), examining learners' general preferences when it comes to learning (Biggs et al., 2001). Researchers in the SAL field distinguish several dimensions of these preferences, such as processing and regulation strategies (Vermunt, 1996). The former are the cognitive activities that students apply when studying. The latter capture the different ways in which students regulate their learning. In assessing these learning strategies, self-report

Likert-type questionnaires mostly relied upon (e.g., Study Process Questionnaire or Inventory of Learning Styles; Biggs et al., 2001; Vermunt, 1998).

Research in the SAL field focuses increasingly on whether and how learning strategies change during the course of higher education (Vanthournout et al., 2011). Examining how these studies are undertaken statistically reveals a strong reliance on comparisons of manifest scale scores over time. For each student, the scores on the items for each scale are averaged at each wave. Subsequently, in studies with two measurement waves, paired-samples *t*-tests are relied upon to compare the means. When more than two measurement waves are involved, repeated measures ANOVA are used.

However, such a straightforward comparison of manifest scale scores over time may be inappropriate when the measurement of the underlying constructs is not equivalent over time: the manifest mean (e.g., the manifest scale scores for the Memorizing scale) depends not only on the latent mean (e.g., being the true Memorizing score at each wave) but on the whole underlying measurement model (Steinmetz, Schmidt, Tina-Booth, Wieczorek, & Schwartz, 2009). Therefore, a longitudinal comparison always hinges upon the assumption of longitudinal measurement invariance (Marsh & Grayson, 1994; Wu et al., 2010). If the ruler does not measure equivalently over time, it is a daunting task to decide whether or not a change in the manifest scale scores is due to actual alterations in learning strategies over time (changes in the latent mean) or due to changes in the measurement over time (Vaillancourt et al., 2003). A measurement can, for example, be age and treatment-sensitive: students having more experience in studying in higher education could interpret learning strategy items differently from novices. Thus, if the assumption of longitudinal measurement invariance is not confirmed, the validity of conclusions stemming from comparisons of manifest scale scores over time could be compromised (Shadish et al., 2002).

Nevertheless, an examination of the measurement model is generally neglected prior to the assessment of change over time (Li, Harmer, & Acock, 1996), perhaps due to a lack of familiarity with the assumption, or with the method of analysis required to verify this. Yet, "[…] whereas it may be reasonable to assume the invariance of these properties over short intervals, this assumption becomes more problematic as time intervals become longer" (Marsh & Grayson, 1994, p.334). Recently, research into changes in learning strategies has increasingly allowed for such longer time intervals (e.g., Donche & Van Petegem, 2009). Thus, the assessment of whether or not the longitudinal measurement assumption holds true, is an evidential lacuna in the learning strategies literature which is becoming increasingly more problematic.

In the methodological literature, testing for measurement invariance across samples (e.g., gender or cross-culturally) is well described (Byrne, 2010). Moreover, though rare in the students' approaches to learning field, numerous applications of multi-sample invariance testing can be found in other social science domains (e.g., Petscher & Huijun, 2008). A large number of these studies rely on data gathered using self-report Likert-type questionnaires. However, the ordinal nature of the data stemming from this is usually ignored by applying a maximum likelihood estimation procedure (e.g., Steinmetz et al., 2009). Studies showcasing measurement invariance testing with a distribution free estimation procedure are scarce. Next to this, measurement invariance testing in longitudinal designs differs from its multi-sample counterpart. Due to the repeated measurements, the responses at different time points are non-independent which, when neglected, can lead to model misspecification (Wu et al., 2010). Moreover, since the number of parameters to be estimated increases rapidly with the number of time points, examining the measurement invariance of all scales together is computationally difficult. Each scale is therefore investigated separately (Vandenberg & Lance, 2000). In sum, the requirements laid on the error terms and the testing procedure differs for longitudinal data compared with multi-sample designs.

In this study, we aim to illustrate longitudinal measurement invariance testing in the SAL domain. By detailing each step in verifying whether or not learning strategy scales measure equivalently over time, we offer a practical guide to longitudinal measurement invariance testing using ordinal data. Moreover, the consequences for the analysis of longitudinal change using scales with varying degrees of measurement invariance are discussed. Therefore, regardless of the research domain tackled here, this study may also be of interest to researchers in other social science fields investigating longitudinal change with self-report Likert-type questionnaires.

3.2 Method

3.2.1 Instrument and sample

As a learning strategy questionnaire, we chose the Inventory of Learning Styles – Short Version (ILS-SV) (Donche & Van Petegem, 2008). This instrument is based on Vermunt's Inventory of Learning Styles (Vermunt, 1996), which was tested cross-culturally (Boyle et al., 2003) and is frequently used in longitudinal research (Vanthournout et al., 2011). The ILS-SV has been validated for first-year University College students, demonstrating the dimensionality of the Vermunt theory, good reliabilities and theoretically sound construct validity (Donche & Van Petegem, 2008).

The ILS-SV questionnaire measures learning strategies consisting of processing and regulation strategies (see Table 3.1). The former are mapped using four scales: Memorizing, Analysing, Critical processing and Relating and structuring. Three scales map regulation strategies: External regulation, Self-regulation and Lack of regulation. All items are scored on a 5-point Likert scale, ranging from (1) 'I never or hardly ever do this', (2) 'I sometimes do this', (3) 'Neutral', (4) 'I often do this' to (5) 'I (almost) always do this'.

Table 3.1: Learning strategies of the ILS-SV questionnaire, scales, number of items, item examples (translated from Dutch) and range of scale reliability

Scales	Items	Item example	Mean inter-item correlation
Processing strategies			
Memorizing	4	I learn definitions by heart and as literally as possible.	.34-.39
Analysing	4	I study each course book chapter point by point and look into each piece separately.	.33-.36
Critical processing	4	I try to understand the interpretations of experts in a critical way.	.32-.39
Relating and structuring	4	I compare conclusions from different teaching modules with each other.	.35-.46
Regulation strategies			
External regulation	5	I study according to the instructions given in the course material.	.20-.27
Self-regulation	4	I use other sources to complement study materials.	.28-.35
Lack of regulation	4	I confirm that I find it difficult to establish whether or not I have sufficiently mastered the course material.	.31-.38

One cohort of students entering a Flemish University College was followed during their three years of higher education. In March of the first academic year (from September to June), all first year students were administered the ILS-SV during scheduled lecture slots. The same cohort had the questionnaire administered again in May of the second and the third year. Though students were not rewarded or given feedback, adequate response rates were obtained each time (73.6%, 67% and 69.8% respectively). Over the three waves, 245 students participated three times. Reliability analysis was conducted using the mean inter-item correlation, since Cronbach alpha values are very sensitive to the number of items (Palant, 2007). At each wave, all scales - containing each 4 to 5 items - met the .2 cut-off for good reliability (see Table 3.1).

Before detailing the measurement invariance testing procedure, we briefly explain the elements in play when assessing the change in learning strategy scales over time. A factor (e.g., the latent concept of Memorizing) is measured at three moments, each time using the same four items (see Figure 3.1; Y_1-Y_4)[2]. The model attempts to predict an individual's score on an item at a certain time (Y_{ijt}).

$$Y_{ijt} = \tau_{jt} + \lambda_{jt}F_{it} + e_{ijt}$$

where i=individual, j=item, t=time

In this prediction, three regression-like elements are key: the intercept (τ_{jt}), the factor loading (λ_{jt}) and the error (e_{ijt}) (see Figure 3.1; Byrne, 2010; Wu et al., 2010). The factor loading (λ_{jt}) represents the increase in Y by one increase in the factor (F_{it}). The intercept (τ_{jt}) can be understood as the value of Y when the latent variable (F_{it}) is zero. Therefore, it reflects the difficulty level or "[…] the ease in getting high manifest scores for a particular measured variable" (Marsh & Grayson, 1994, p. 336).

However, in our case, the items are ordinal. Therefore, there is not one intercept, but several thresholds. With a five-point Likert scale, there are four thresholds (the number of scale points – 1) (Metha et al., 2004). For example, $\tau_{3;\ \text{time 2};\ \text{threshold 1}}$ expresses for item 3 at time 2 the difficulty level of scoring *I sometimes do this* (Likert point 2) compared to *I never or hardly ever do this* (Likert point 1) when the latent variable (F_{it}) is zero.

---

[2] Figure 3.1 depicts the situation for all scales except the External regulation scale, having not 4 but 5 items (see Table 3.1).

The third element in the equation is the measurement error (e_{ijt}). Due to the data's longitudinal nature, it is plausible that errors pertaining to the same item (e.g., e_{11}, e_{12} and e_{13}, see Figure 3.1) correlate over time (Vaillancourt et al., 2003). To prevent model misspecification, three item covariances are estimated per item (e.g., for Y_1: e_{11}-e_{12}, e_{11}-e_{13} and e_{12}-e_{13})[3] (Wu et al., 2010).

To assess change, the scores on the four items (Y's) are usually averaged for each student per wave. Subsequently, manifest scale scores are compared over time. Conclusions are then drawn in terms of the underlying latent factors (F's) (e.g., Memorizing decreases during higher education). Yet, change in item scores over time (ΔY) can only be attributed to change in this latent factor (ΔF_{it}) when the other elements in the equation remain invariant over time (Byrne, 2010; Marsh & Grayson, 1994). However, due to the correlation between errors over time, and contrary to multi-group comparisons, error invariance is not expected in longitudinal measurement invariance testing (Wu et al., 2010). The longitudinal measurement invariance analysis of ordinal data thus consists of two elements - the invariance of factor loadings (λ's) and of thresholds (τ's).

[3] These error covariances have not been drawn in Figure 3.1 in order not to complicate the figure excessively.

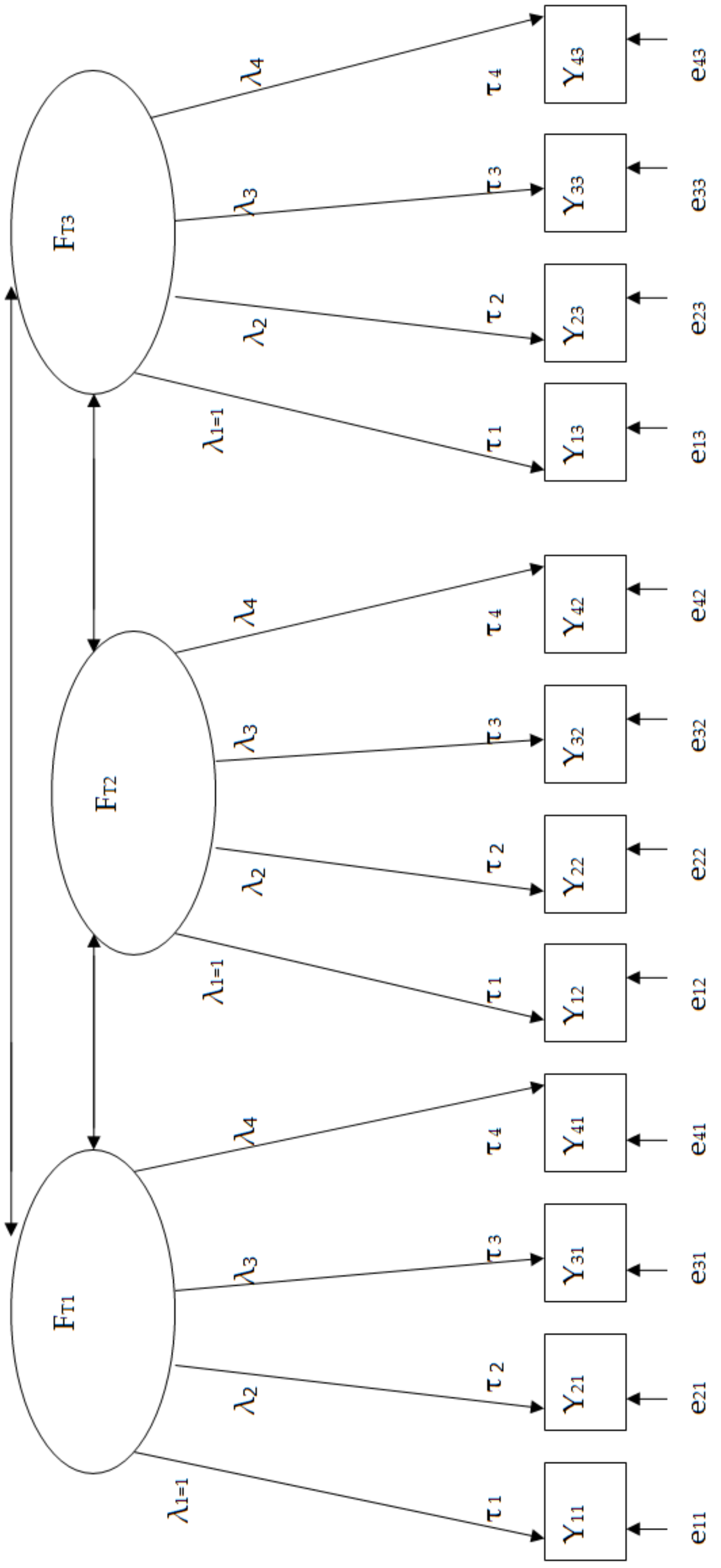


Figure 3.1: Longitudinal measurement model
(based on Wu et al., 2010)

3.2.2 Procedure for longitudinal measurement invariance testing

In testing whether the measurement invariance hypothesis holds, successively more constrained models are estimated for each scale (see Figure 3.2; Muthén & Muthén, 2010). Due to the data's ordinal nature, the use of the maximum likelihood estimation procedure could not be justified. Therefore, a distribution-free estimation procedure, the weighted least squares means-variance (WLSMV), was employed in Mplus 6.1[4] (Metha et al., 2004; Muthén & Muthén, 2009b).

First, a baseline model is estimated, testing whether for each scale a unidimensional model holds at each measurement point (Vandenberg & Lance, 2000). To evaluate this, neither factor loadings nor thresholds are constrained to be equal over time, while the error covariances are included. Subsequently, an adequate fit is suggested by a Comparative Fit Index (CFI) close to .95 (Hu & Bentler, 1999) and an RMSEA up to .08 (Byrne, 2010).

[4] Mplus Version 6.1 syntaxes are provided in 3.5.

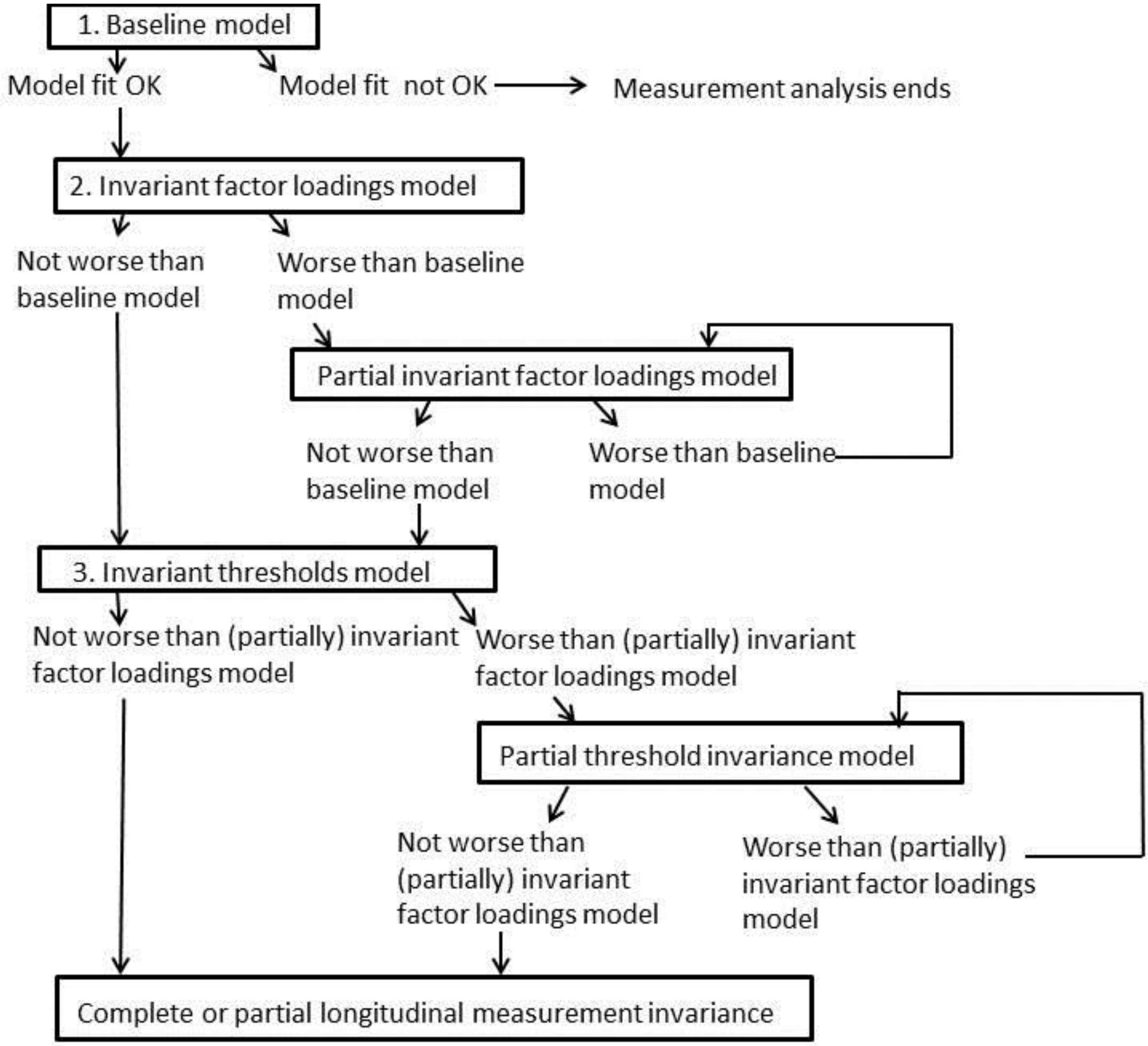


Figure 3.2: Flowchart for longitudinal measurement invariance testing

In the second model, for each item, the factor loadings (λ's) are constrained to be equal over time (e.g., $\lambda 2$ at time 1= $\lambda 2$ at time 2= $\lambda 2$ at time 3; Wu et al., 2010). Subsequently, the hypothesis of invariance is evaluated by comparing the model fit of the more restricted invariant factor loadings model, to the less restricted baseline model. To test this, the Chi-square difference test ($\Delta\chi^2$) and the change in CFI (ΔCFI) criterion are relied upon (Byrne, 2010; Vandenberg & Lance, 2000). For the former, the hypothesis of equal factor loadings over time is rejected when the Chi-square difference test ($\Delta\chi^2$) has a probability lower than 0.05[5]. For the latter, a decrease in CFI by 0.01 or more suggests that the invariance hypothesis should be rejected[6] (Chueng & Rensvold, 2002). Failure to reject the hypothesis is interpreted as evidence that an increase of 1 in the factor score (F_{it}) procures the same increase (λ_2) in the item (Y_2) at each wave. If the hypothesis of equal factor loadings is rejected, this signifies that (at least) one of the items is more or less closely related to the underlying construct at one time rather than at the other (Cooke, Kosson, & Michie, 2001).

In this situation, additional models are warranted to identify the source(s) of the lack of equivalence. High values on the modification indices (Mod. Ind.) and the expected parameter change (EPC) suggest that the constraint on the factor loading needs to be freed (Muthén & Muthén, 2009b). If such a partial factor loadings invariance model produces a non-significant loss of fit compared to the baseline model (*p* of $\Delta\chi^2$>.05; ΔCFI>-.01), all factor loadings can be assumed to be equal besides the one freely estimated. If the model fit is still worse in relation to the baseline model, the above procedure is repeated (see Figure 3.2).

---

[5] Due to the WLSMV estimator used here, the change in Chi² and degrees of freedom cannot be calculated in a straightforward fashion. "The difference in chi-square values for two nested models using the […] WLSMV chi-square values is not distributed as chi-square" (Muthén & Muthén, 2009b, p. 501). Therefore, a scaling correction (DIFFTEST function) is relied upon, of which only the *p*-value should be interpreted.

[6] The $\Delta\chi^2$ and ΔCFI may sometimes suggest different conclusions. Clear rules on how to proceed in such situation are lacking (Byrne, 2010), though in large samples, the ΔCFI may be more credible than the $\Delta\chi^2$ (Meade, Johnson, & Braddy, 2008). In other cases, researchers can opt to describe the conclusions of both approaches or choose one over the other (Byrne, 2010), based upon the admissibility of the solution and examination of the MI.

Next, equality constraints on the thresholds (τ's) are added. For each item, it is verified whether or not the difficulty level of going, for example, from *I often do this* to *I (almost) always do this*, remains constant over time (e.g., $\tau_{2\ \text{time 1; threshold 4}} = \tau_{2\ \text{time 2; threshold 4}} = \tau_{2\ \text{time 3; threshold 4}}$). A non-significant loss of fit of the invariant thresholds model compared to the (partially) invariant factor loadings model (p of $\Delta\chi^2>.05$; ΔCFI>-.01), suggests that the thresholds can be assumed to be equally difficult over time. Rejection of the equal thresholds hypothesis indicates that the difficulty level for (at least) one threshold varies over time (Metha et al., 2004). By freeing the constraint on the threshold causing most trouble according to the Mod.Ind. and EPC, a partial threshold invariance model is estimated.

How many factor loadings and thresholds can be freed without jeopardising future longitudinal analysis constitutes a debate in the literature (Byrne, 2010; Marsh & Grayson, 1994). Differences in factor loadings are, however, perceived to be more serious in relation to bias than differences in thresholds (Cooke et al., 2001). Therefore, we judge complete invariance of factor loadings as a necessary condition for longitudinal analysis. Concerning the number of unequal thresholds which are tolerable, a minimum of two items for which all thresholds are invariant is suggested (Steinmetz et al., 2009).

3.3 Results

3.3.1 Processing strategies

The baseline model of the Memorizing scale showed adequate fit (see Table 3.2), indicating that the Memorizing scale is unidimensional at each measurement wave. In testing the invariance of the factor loadings, a non-significant loss of fit with respect to the unconstrained baseline model was obtained ($\Delta\chi^2$=2,277, Δdf=6, p=.89; ΔCFI=.008). The discrepancy between the invariant thresholds model and the invariant factor loadings model also satisfied the minimum criteria for invariance ($\Delta\chi^2$=13,378, Δdf=22, p=.92; ΔCFI=.003). Complete longitudinal measurement invariance can thus be assumed for the Memorizing scale.

For the Analysing scale, the baseline model also shows adequate model fit and constraining the factor loadings does not alter the model fit significantly ($\Delta\chi^2$=4,115, Δdf=6, p=.66; ΔCFI=.008). However, the invariant thresholds hypothesis is rejected ($\Delta\chi^2$=40,574, Δdf=22, p<.001; ΔCFI=-.015). The second threshold (going from *I sometimes do this* to *neutral*) of the item "I study each course book chapter point by point and look into each piece separately" is less difficult at the third wave (Mod.Ind.=6.836, EPC=-.180). Relaxing the constraint on this threshold did not improve model fit sufficiently ($\Delta\chi^2$=32,320, Δdf=21, p=.054; ΔCFI=-.01). A re-examination of the Mod.Ind. pointed anew to the same item: the difficulty of answering *I (almost) always do this* is higher at the first wave (Mod.Ind.=5.732, EPC=.180). Allowing this threshold to be freely estimated provided a model that was statistically indistinguishable from the equal factor loadings model ($\Delta\chi^2$=25,599, Δdf=20, p=.17; ΔCFI=-.006). The results for the Analysing scale thus suggested factor loadings invariance and the equality of all but two thresholds pertaining to the same item.

Concerning Critical processing, the baseline model suggests an adequate model fit. The hypothesis of invariant factor loadings was not rejected ($\Delta\chi^2$=9,278, Δdf=6, p=.16; ΔCFI=-.002) and constraining the thresholds did not decrease model fit ($\Delta\chi^2$=22,637, Δdf=22, p=.42; ΔCFI=-.003). The results for the Relating and structuring scale paint a similar picture. Both the factor loadings and the thresholds can be presumed to be equal over time (respectively, $\Delta\chi^2$=8,912, Δdf=6, p=.18; ΔCFI=.002 and $\Delta\chi^2$=22,429, Δdf=22, p=.44; ΔCFI=-.003). Consequently, for the scales Critical processing and Relating and structuring, the results indicate complete longitudinal invariance of factor loadings and thresholds.

3.3.2 Regulation strategies

The fit of the baseline model of the External regulation strategy scale suggests that the unidimensionality of the scale holds over the three waves. Constraining factor loadings did not produce a significant worsening of fit ($\Delta\chi^2$=5,342, Δdf=8, p=.72; ΔCFI=.011), while the invariant thresholds model did ($\Delta\chi^2$=47,944, Δdf=28, p<.05; ΔCFI=-.016). The item "I study according to the instructions given in the course material" failed to reveal invariance at the second measurement wave for the fourth threshold (Mod.Ind.=10.008, EPC=-.536). It was less difficult to answer *I (almost) always do this* in the second year. Results for the External regulation scale thus suggest invariance over time of the factor loadings and all but one threshold.

Table 3.2: Results from the measurement invariance tests for processing and regulation strategy scales

	Model description	χ^2	df	CFI	RMSEA	$\Delta\chi^2$	Δdf	p	ΔCFI
Memorizing	Baseline	54,670	39	.989	.040				
	Invariant loadings	49,493	45	.997	.020	2,277	6	.893	.008
	Invariant thresholds	66,127	67	1.000	.000	13,378	22	.922	.003
Analysing	Baseline	82,231	39	.967	.067				
	Invariant loadings	77,376	45	.975	.054	4,115	6	.661	.008
	Invariant thresholds	118,445	67	.960	.056	40,574	22	***	-.015
	Partial threshold invariance	111,011	66	.965	.053	32,320	21	.054	-.010
	Partial threshold invariance	105,520	65	.969	.050	25,599	20	.179	-.006
Critical processing	Baseline	47,445	39	.994	.030				
	Invariant loadings	56,875	45	.992	.033	9,278	6	.158	-.002
	Invariant thresholds	82,600	67	.989	.031	22,637	22	.422	-.003
Relating and structuring	Baseline	64,925	39	.986	.052				
	Invariant loadings	67,655	45	.988	.045	8,912	6	.179	.002
	Invariant thresholds	94,712	67	.985	.041	22,429	22	.435	-.003
External regulation	Baseline	126,532	72	.950	.056				
	Invariant loadings	122,224	80	.961	.046	5,342	8	.72	.011
	Invariant thresholds	168,162	108	.945	.048	47,944	28	*	-.016
	Partial threshold invariance	158,145	107	.953	.044	34,424	27	.154	-.008
Self-regulation	Baseline	50,903	39	.991	.035				
	Invariant loadings	74,066	45	.998	.014	1,809	6	.936	.007
	Invariant thresholds	67,851	65	.998	.013	19,990	20	.459	.000

(table continues)

Table 3.2 (continued): Results from the measurement invariance tests for processing and regulation

	Model description	χ^2	df	CFI	RMSEA	$\Delta\chi^2$	Δdf	p	ΔCFI
Lack of regulation	Baseline	90,467	39	.969	.073				
	Invariant loadings	73,890	45	.983	.051	2,685	6	.847	.014
	Invariant thresholds	103,456	67	.978	.047	26,346	22	.237	-.005

*p <.05; **p <.01; ***p <.001

Concerning the second scale, Self-regulation, the baseline model shows adequate fit and the hypothesis of factor loading invariance is not rejected ($\Delta\chi^2$=1,809, Δdf=6, p=.94; ΔCFI=.007). Constraining thresholds over time however, proves problematic for the item "I use other sources to complement study materials". At both the second and the third wave, no students answered *I (almost) always do this*. At the first measurement wave, this answered is checked by less than 1% of the students. The invariant thresholds model (not estimating the two absent thresholds) fitted the data as well as the invariant factor loadings model ($\Delta\chi^2$=19,990, Δdf=20, p=.46; ΔCFI=.000), indicating that the measurement of Self-regulation can be assumed equivalent over time.

Lastly, for the Lack of regulation scale, the model fit for the baseline model suggests unidimensionality, and the discrepancy between the invariant factor loadings model and the baseline model satisfied the minimum criteria for invariance ($\Delta\chi^2$=2,685, Δdf=6, p=.85; ΔCFI=.014). Moreover, constraining the thresholds over time produces a non-significant loss of fit ($\Delta\chi^2$=26,346, Δdf=22, p=.24; ΔCFI=-.005). It is therefore concluded that the Lack of regulation scale measures equivalently over time.

3.4 Discussion

In the SAL field, a growing number of studies have examined whether and how learning strategies evolve over the course of higher education. To assess this, comparisons of manifest scale scores over time by means of t-tests and repeated measures ANOVA are used. An often overlooked assumption of these techniques is that the ruler needs to measure equivalently at each wave. Taking the case of the learning strategies scales, the current study therefore illustrates the longitudinal measurement invariance testing procedure with ordinal Likert-type data.

The results confirm at least partial measurement invariance for the four processing and the three regulation scales of the ILS-SV. All factor loadings pertaining to the scales proved invariant over measurement waves as well as did at least all thresholds belonging to two items. This is a promising result since significance testing on found mean differences is only permitted if this minimal degree of partial invariance is confirmed (Steinmetz et al., 2009). However, which analytical technique is most adequate, depends on the degree of invariance of a scale.

For five learning strategies scales complete measurement invariance was confirmed, ensuring a comparable definition of the latent construct over time. In this situation, traditional statistical comparison procedures such as *t*-tests or repeated measures ANOVA on manifest scale scores are non-problematic (Steinmetz et al., 2009; Vandenberg & Lance, 2000). For the External regulation and Analysing scale, respectively one and two thresholds failed to reveal equivalence over measurement moments. These variances can seriously hamper the comparison of manifest scale scores, since it is difficult to disentangle genuine changes in the underlying latent variable from nuisance due to shifts in the difficulty level (Steinmetz et al., 2009). Therefore, it is suggested that researchers refrain from traditional statistical procedures and explicitly model the small number of variations via a structural equation modelling procedure such as a multiple indicator latent growth model (Marsh & Grayson, 1994; Vandenberg & Lance, 2000).

3.4.1 Limitations and future studies

Certain limitations of the current study suggest additional avenues for future research. Firstly, there are different techniques to assess measurement invariance. Here, the approach based upon confirmatory factor analysis was used, while Fidalgo and Scalon (2010), for example, relied upon a IRT-based differential item functioning technique. It would be interesting to assess the impact of these different techniques in longitudinal measurement invariance testing. Second, when equivalence of the measurement model is established, the structural invariance can be assessed. For example, the evolution of the correlation between scales can be substantively relevant (Chueng & Rensvold, 2002). In the SAL field it is, for example, theoretically viable for scales to be differently related over time. Third, the results from this study cannot be generalized to other educational contexts, cultures, learning strategy questionnaires or samples. Comparable to reliability, longitudinal measurement invariance should be assessed anew in each specific sample (Guttmannova, Szanyi, & Cali, 2008).

The limitations of the present study notwithstanding, the results provide apparent support for the need for longitudinal measurement equivalence testing. As was succinctly stated by Wu and colleagues (2010) "[…] establishing temporal measurement invariance is the prerequisite for analyzing change" (p. 126). We therefore hope to have provided a clear illustration of the longitudinal measurement invariance testing procedure in the case of ordinal data stemming from Likert-type questionnaires.

3.5 Mplus syntaxes

3.5.1 Baseline model

```
TITLE: factor 1 baseline
DATA:
   FILE IS C:\Desktop\factorone.txt;
   NOBSERVATIONS = 245;
VARIABLE:
   NAMES ARE u11 u21 u31 u41
            u12 u22 u32 u42
            u13 u23 u33 u43;
   CATEGORICAL ARE u11 u21 u31 u41
            u12 u22 u32 u42
            u13 u23 u33 u43;
MODEL: f1 BY u11-u41;
f2 BY u12-u42;
f3 BY u13-u43;
!correlated residual errors
u11 with u12 u13;
u12 with u13;
u21 with u22 u23;
u22 with u23;
u31 with u32 u33;
u32 with u33;
u41 with u42 u43;
u42 with u43;
OUTPUT:
STANDARDIZED TECH4;
MODINDICES(0);
SAVEDATA: FILE is factorone.dat;
DIFFTEST IS deriv.dat;
SAVE=fscores;
```

3.5.2 Invariant factor loadings model

```
ANALYSIS: DIFFTEST IS deriv.dat;
MODEL: f1 BY u11
```

```
 u21-u41 (1-3);
f2 BY u12
u22-u42 (1-3);
f3 BY u13
u23-u43 (1-3);
.
SAVEDATA: FILE is factorone.dat;
.
```

3.5.3 Partial invariant factor loadings model

! at the third wave, the factor loading of the second item (u23) is not constrained equal to the !factor loadings at time one (u21) and two (u22)

```
MODEL: f1 BY u11
 u21-u41 (1-3);
f2 BY u12
u22-u42 (1-3);
f3 BY u13
u23-u43 (2-3);
```

3.5.4 Invariant thresholds model

```
.
MODEL: f1 BY u11
 u21-u41 (1-3);
f2 BY u12
u22-u42 (1-3);
f3 BY u13
u23-u43 (1-3);
[u11$1 u12$1 u13$1] (4);
[u11$2 u12$2 u13$2] (5);
[u11$3 u12$3 u13$3] (6);
[u11$4 u12$4 u13$4] (7);
[u21$1 u22$1 u23$1] (8);
[u21$2 u22$2 u23$2] (9);
[u21$3 u22$3 u23$3] (10);
[u21$4 u22$4 u23$4] (11);
[u31$1 u32$1 u33$1] (12);
```

```
 [u31$2 u32$2 u33$2] (13);
 [u31$3 u32$3 u33$3] (14);
 [u31$4 u32$4 u33$4] (15);
 [u41$1 u42$1 u43$1] (16);
 [u41$2 u42$2 u43$2] (17);
 [u41$3 u42$3 u43$3] (18);
 [u41$4 u42$4 u43$4] (19);
{u11-u41@1 u12-u43};
[f1@0 f2 f3];
.
```

3.5.5 Partial thresholds invariance model

```
! the third threshold for the second item is estimated freely at the first wave (u21$3), while !constrained equal at the second and third wave (u22$3 u23$3)
.
[        u22$3 u23$3] (10);
```

4. The growth trend in learning strategies during higher education

This chapter is based on Coertjens, L., Donche, V., De Maeyer, S., Vanthournout, G., & Van Petegem, P. (2013). Modeling change in learning strategies throughout higher education: A multi-indicator latent growth perspective. *PLoS ONE, 8*(7), e67854. doi: 10.1371/journal.pone.0067854

The change in learning strategies during higher education is an important topic of research in the Students' Approaches to Learning field. Although the studies on this topic are increasingly longitudinal, analyses have continued to rely primarily on traditional statistical methods. The present research is innovative in the way it uses a multi-indicator latent growth analysis in order to more accurately estimate the general and differential development in learning strategy scales. Moreover, the predictive strength of the latent growth models are estimated. The sample consists of one cohort of Flemish University College students, 245 of whom participated in the three measurement waves by filling out the processing and regulation strategies scales of the Inventory of Learning Styles – Short Versions. Independent-samples t-tests revealed that the longitudinal group is a non-random subset of students starting University College. For each scale, a multi-indicator latent growth model is estimated using Mplus 6.1. Results suggest that, on average, during higher education, students persisting in their studies in a non-delayed manner seem to shift towards high-quality learning and away from undirected and surface-oriented learning. Moreover, students from the longitudinal group are found to vary in their initial levels, while, unexpectedly, not in their change over time. Although the growth models fit the data well, significant residual variances in the latent factors remain.

4.1 Introduction

How students go about their learning has been one of the core interests of educational researchers. This topic has been investigated with regard to many stages of formal education (Rozendaal, Minnaert, & Boekaerts, 2003; van Bragt, Bakx, Van der Sanden, & Croon, 2007), one important one being students in higher education. This issue has been looked at from different angles, with the Students' Approaches to Learning (SAL) being a particularly long-lived one (Richardson, 2011).

The SAL field is comprised of different theories describing students' varying preferences for learning strategies (Coffield et al., 2004). Of the questionnaires associated with the SAL theories, Richardson (2004) distinguishes the two most frequently used with campus-based students. Firstly, Biggs and colleagues' Study Process Questionnaire (SPQ, Biggs et al., 2001) discerns two main ways to go about learning. The deep approach can be conceptualised as a combination of aiming for understanding and using strategies to create meaning; for example relating aspects of the content with one another. The surface approach can be understood as using memorizing techniques (learning by heart) with the aim of passing the course or task. Initially, a third approach was discerned, labelled achieving or strategic approach and capturing students' strategies to maximise their grades. This factor was however found to load on the deep approach or, sometimes, on the surface approach (Biggs et al., 2001; Kember & Leung, 1998). Therefore, in a revised version of the SPQ, only the first two approaches were withheld (Biggs et al., 2001). Secondly, Entwistle and colleagues (Entwistle & Ramsden, 1983) developed the Approaches to Study Inventory (ASI) containing three orientations to learning. The meaning orientation and reproducing orientation can be conceptually linked to the deep and surface approach, respectively. The achieving orientation can be thought of in line with the achieving or strategic approach.

To these two questionnaires, Fox et al. (2010) and Vanthournout et al. (2011) add the Vermunt's Inventory of Learning Styles (ILS) Questionnaire, which maps four elements: (1) processing strategies; (2) regulation strategies; (3) conceptions of learning and; (4) learning orientations (Vermunt & Vermetten, 2004). The first two elements --processing strategies and regulation strategies-- are sometimes subsumed under the concept of learning strategies. These learning strategies can be related to the learning approaches or orientations described above. Processing strategies refer to those cognitive strategies that students apply whilst studying. In the ILS, two cognitive processing strategies map deep processing: critical processing and relating and structuring. Stepwise or surface processing is captured by the memorizing and analysing scale. Regulation strategies are those meta-cognitive activities students use to direct their learning process, such as planning or testing oneself. The ILS incorporates three such strategies. The self-regulation scale refers to directing the learning process oneself. External regulation captures the degree to which students seek guidance by the teacher or by the learning material. The lack of regulation scale expresses whether students are undirected in their learning, i.e. they do not steer themselves nor follow their teachers' guidance.

The different theories on learning strategies assume linkage with academic performance. Concerning the SPQ, deep and surface processing strategies are expected to lead to higher or lower achievement, respectively (Marton & Säljö, 1976). A recent review study confirmed this: the correlation between grade point average (GPA) and deep and surface approaches is small positive and small negative, respectively (Richardson, Abraham, & Bond, 2012). Using the ASI or an adaptation of it (Revised ASI, RASI, or Approaches to Study Skills Inventory for

Students, ASSIST), a number of studies found small positive correlations between meaning orientation and grade (Diseth, 2007; Sadler-Smith, 1996; Sadler-Smith & Tsang, 1998), though the study by Provost and Bond (1997) did not detect an association. The reproducing orientation was in most studies found to have a small negative association with achievement (Diseth, 2007; Provost & Bond, 1997; Sadler-Smith & Tsang, 1998).

Regarding the ILS, empirical research shows positive and weak-to-moderate correlations between deep processing and academic achievement (Boyle et al., 2003; Donche & Van Petegem, 2011; Vanthournout et al., 2012; Vermunt, 2005). For surface processing, the memorizing strategy is unrelated to performance whilst the evidence for the analysing strategy is mixed between a positive association (Boyle et al., 2003; Donche & Van Petegem, 2011) and absence of a correlation (Vanthournout et al., 2012; Vermunt, 2005). For the regulation strategies, the evidence on self-regulation and external relation is unequivocal whilst the unregulated learning strategy is repeatedly found to be related to lower academic achievement (Donche & Van Petegem, 2011; Vanthournout et al., 2012; Vermunt, 2005).

Though learning strategies are clearly not the sole predictor of academic achievement in general, for the three learning strategy questionnaires (SPQ, ASI and ILS), deep- or meaning-oriented learning has a small positive correlation with academic achievement. For surface/reproducing learning, in most studies, a small negative association with academic achievement is found. Last, unregulated learning is associated with lower academic performance.

The association between learning strategies and learning outcomes is not only important during higher education but also afterwards, for example in the working context. It is presumed that deep learning during higher education is linked to being a reflective and adaptive practitioner, who participates in lifelong learning (Kyndt, Dochy, Cascallar, & Struyven, 2011; Reid et al., 2005). Though more research is clearly needed to test this presumption, Hoeksema and colleagues (1997) detected that deep learning correlates with career success whilst surface learning was found to hamper it. Self-directed learning was also found to be positively associated with the amount of work-related learning (Gijbels, Raemdonck, & Vervecken, 2010).

Next to literature on the link between learning strategies, achievement and lifelong learning, numerous studies in the SAL field debate whether learning strategies should be conceptualised as a trait or a state (Richardson, 2011; Vermetten, Lodewijks, & Vermunt, 1999a). Though some view them to be fixed personality-like characteristics (Messick, 1996), there is a body of literature on the influence of person-related factors, such as age and motivation (for a review, see Baeten et al., 2010), supporting the view of learning strategies as a state. Next to this, the influence of contextual factors - for example, elements in the learning environment such as teachers' approaches on students' learning strategies or assessment - has been described (Baeten et al., 2010; Entwistle et al., 2003; Trigwell, Prosser, & Waterhouse, 1999). Moreover, a number of studies have detected change in learning strategies over time (Phan, 2011; Reid et al., 2005; Vanthournout, 2011).

4.1.1 Change in learning strategies

Change in learning strategies over time has been on the one hand assessed using cross-sectional designs. Relying on the SPQ, both Gow & Kember (1990) and Biggs (1987) found that students in higher years at university scored lower on both the deep and achieving approach. No significant differences between the different years was noted for the surface approach. Using the RASI, Richardson (2006) detected lower levels for the deep approach in higher years, while they were higher for the surface approach. There was no significant difference between the years concerning the strategic approach.

It can be argued that found effects in cross-sectional designs could alternatively be explained by varying group composition between the years. Therefore, researchers have assessed changes in learning strategies by using repeated measurements of for example the SPQ, ASI or ILS with the same students (Reid et al., 2005; Van der Veken et al., 2009; Zeegers, 2001). Historically, this research has relied on pre-test post-test designs over a short interval of time (Dart & Clarke, 1991; Fox et al., 2010; Vermetten et al., 1999b; Volet, Renshaw, & Tietzel, 1994; Watkins & Hattie, 1985). Recently, more than two measurement waves have been taken into account and longer time intervals have been allowed for (Gordon & Debus, 2002; Zeegers, 2001), which comply with the criteria for longitudinal research as put forward by Singer and Willet (2003). For each of the three theoretical frameworks, we will briefly discuss the findings of these longitudinal studies.

Relying on Biggs and colleagues' framework, Phan (2011), Gordon and Debus (2002) and Jackling (2005) noted an increase in deep processing over time. On the other hand, Zeegers (2001) concluded that it remained constant. These results clearly contradict the findings in the cross-sectional studies mentioned earlier, in which the deep approach was found to be lower for students in later years of higher education. Concerning surface processing, Gordon and Debus (2002) detected a decreasing reliance, while Jackling (2005) and Zeegers (2001) found a constant trend. The change in the achieving approach was only investigated by Gordon and Debus (2002), who concluded that it remained constant over time. Using the ASSIST (Tait et al., 1998), Reid and colleagues (2005) noted a decrease in both the deep and the strategic approach over the first year of medical training, while the surface approach did not alter. Over the second year, the surface and deep approach remained stable, while the strategic approach continued to decrease.

Five longitudinal studies use the Vermunt framework and the ILS to map changes throughout higher education (Busato et al., 1998; Donche et al., 2010; Severiens et al., 2001; Van der Veken et al., 2009; Vanthournout, 2011). Concerning meaning-oriented learning (Busato et al., 1998) or deep processing strategies (Donche et al., 2010; Severiens et al., 2001; Van der Veken et al., 2009) an increase was found. However, Vanthournout (2011) concluded that only the relating and structuring scale increases, while the degree of critical processing remains constant. Stepwise processing and its subscale analysing was found to remain constant over time (Busato et al., 1998; Donche et al., 2010; Severiens et al., 2001; Vanthournout, 2011), while the degree of memorizing decreased over time (Donche et al., 2010) or showed a quadratic trend with a rise after an initial decrease (Vanthournout, 2011). Concerning the regulation strategies, self-regulation was found to increase (Donche et al., 2010; Severiens et al., 2001; Vanthournout, 2011), or to remain constant (Van der Veken et al., 2009). External regulation, on the other hand, decreased (Donche et al., 2010; Severiens et al., 2001; Vanthournout, 2011). Lastly, undirected learning was found to remain constant (Busato et al., 1998; Severiens et al., 2001; Van der Veken et al., 2009) or decrease over time (Donche et al., 2010; Vanthournout, 2011).

4.1.2 Restraints of the statistical techniques used to assess change in learning strategies

An examination of how longitudinal studies within the SAL field are undertaken statistically reveals a strong reliance on comparisons of manifest scale scores over time. Per student, the scores on the items for each scale are averaged at each wave. Subsequently, repeated measures (M)ANOVA are relied upon to compare the mean factor scores over time (Jackling, 2005; Van der Veken et al., 2009). This type of analysis discerns whether students, *on average,* increase or decrease in terms of a particular learning strategy scale.

When average growth is estimated by comparing manifest scale scores over time, two important elements are overlooked. First, it remains veiled whether students follow a comparable growth trajectory or whether they differ in their growth over time. Studies on this differential change in learning strategies are explicitly called for in literature (Nienemin et al., 2004). However, evidence concerning the differential growth in learning strategies during higher education is scarce. At present, only two studies have looked at this in an in-depth fashion (Phan, 2011; Vanthournout, 2011).

Second, all studies on the changes in learning strategies over time have relied upon manifest scale scores to draw conclusions with regard to latent factors (e.g. deep learning decreases during higher education). Hereby, a number of measurement issues are overlooked. By using manifest scale scores, the measurement error associated with learning strategy items, the ordinal nature of Likert scale items and the assumption of measurement invariance are ignored. In what follows, we will first detail the prior findings concerning differential growth. Next, the limitations of manifest scale scores are discussed.

4.1.3 Differential change in learning strategies

A number of studies have investigated the influence of initial learning strategies over a short period of time (Baeten et al., 2010). At the course-level, Wilson and Fowler (Wilson & Fowler, 2005) detected that students scoring high on the deep approach did not vary in their reliance on this strategy between a conventional and action learning course. One the other hand, students judged 'typically surface' reported greater use of deep learning strategies in the action learning course. Studies by Gijbels and colleagues (2009) and Vanthournout (2011) confirm that initial learning strategies influence the change in these strategies during a course.

Four studies have looked at the differential evolution in learning strategies over longer periods of time (Donche et al., 2010; Nienemin et al., 2004; Phan, 2011; Vanthournout, 2011). Two studies performed preliminary analysis on subgroups of students. Donche et al. (2010) examined whether students' changes in learning strategy scales during their time in higher education was dependent on the learning profile upon entry into higher education. Relying on cluster analysis and subsequently paired-samples t-tests per cluster, the authors concluded that there is some evidence that subgroups of students develop in different ways. Nienemin and colleagues (2004) detected small differences in the development of learning strategies among students scoring below average, compared to their peers scoring above average.

The two other studies have used more advanced analysis to model differential evolution explicitly. Relying on a multilevel model, Vanthournout (2011) found a differential evolution in change over time for the critical processing, self-regulation, analysing and external regulation scales. For the last two scales, this correlated negatively with students' initial level: students scoring higher at the start of higher education tended to decrease their reliance, while those scoring lower initially tended to increase it. Using a latent growth model, Phan (2011) concluded upon a comparable differential growth in deep processing: students scoring lower on deep processing at the start of their undergraduate program were found to increase their reliance more rapidly.

4.1.4 Limitations of manifest scale scores

Up to present, all studies on the change in learning strategies over time have used manifest scale scores. These manifest scale scores ignore however three measurement issues: measurement error, the ordinal nature of Likert scale items and measurement invariance over time. By relying on multi-indicator latent growth (MILG) analysis, average as well as differential growth can be estimated while taking each of these measurement issues adequately into account.

First, as confirmatory factor analyses on learning strategy questionnaires confirm, items do not perfectly measure a certain concept, but have measurement error (Boyle et al., 2003). Such measurement error can be explicitly modelled in MILG analysis (Byrne, 2010), thereby estimating a latent mean per measurement wave, which can be subsequently compared over time (Metha et al., 2004; Wu et al., 2010).

Second, the three measurement instruments described (SPQ, ASI & ILS) use Likert scales. By averaging the scores on the items for one learning strategy scale and applying a repeated-measures (M)ANOVA, it is implicitly assumed that the manifest scale scores are continuous. Yet, whether this assumption holds true, is debatable for Likert scales (Metha et al., 2004; Muthén & Muthén, 2010). In MILG analysis, the ordinal nature of these scales can again be explicitly accounted for, allowing us to '… make inferences about change on an interval metric when all we have are data on ordinal metric' (Metha et al., 2004, p. 304).

Third, the comparison of learning strategy factor scores over time is based upon the assumption of measurement invariance (Marsh & Grayson, 1994; Muthén & Muthén, 2009b). What the learning strategy questionnaire actually measures needs to be equivalent at each wave to allow for comparisons over time (Stoel et al., 2006). This assumption applies to (M)ANOVA's, multilevel models and MILG analysis alike. (M)ANOVA's and multilevel models do not provide the opportunity to falsify this assumption, given that they rely on the manifest scale scores. MILG analysis on the other hand models growth in the latent scale scores. Together with the growth trend, the factor structure (factor loadings and item difficulty) is thus estimated. This allows verification of whether factor loadings and item difficulty remains constant over time (i.e. measurement invariance).

Furthermore, MILG analysis allows for testing how good a growth trajectory predicts the true, latent change in a learning strategy scale. Not only fit indices, but also R^2 parameters, indicate the predictive power of the growth trajectory (Voelkle et al., 2006). Moreover, residual variances suggest whether additional predictors are needed to adequately estimate the change in learning strategy scales over time.

4.1.5 The current research

The current research further analyses the data of the Donche et al. (2010) study, in which repeated measures ANOVA were used. By accounting for measurement error, measurement variance and the ordinal nature of the data, we aim to answer the following research questions (RQ): how do student's processing and regulation strategies develop on average? (RQ1), is there differential growth in learning strategies? (RQ2) and, how much variance in the latent factors is explained by the growth factors? (RQ3)

4.2 Materials and Method

4.2.1 Ethics Statement

For research in higher education, ethics approval and written consent is not required by Belgian law. The Law on Experiments on Humans (7th May 2004) obliges ethics approval and consent for an experiment, whereby 'experiment' is defined as "each study or research in which human persons are involved with the goal of developing appropriate knowledge for the performance of health professions" ("elke op de menselijke persoon uitgevoerde proef, studie of onderzoek, met het oog op de ontwikkeling van de kennis eigen aan de uitoefening van de gezondheidszorgberoepen", 2004050732/N, Article 2, paragraph 11). The current research is not related to the performance of health professions and is therefore implicitly exempt from ethics approval and written consent. We underline that participation was at each wave on a voluntary basis and that the students, who were all adults, could stop their participation at any moment. There was no penalty for students who chose not to participate, nor were they rewarded for participation with, for example, student counselling regarding learning strategies. Confidentiality of the results was guaranteed by the research team.

4.2.2 Measurement

Learning strategies are investigated by focusing on two malleable components of the learning patterns of the Vermunt framework: cognitive processing and regulation activities (Vermunt, 1998). The scales used in this study stem from the 'Inventory of Learning Styles – Short Version' (ILS-SV), which has been validated for first-year Flemish University College students (Donche & Van Petegem, 2008). Processing strategies can be viewed as the cognitive activities a student applies whilst studying. In the ILS-SV, four scales for cognitive processing strategies are distinguished: memorizing, analysing, critical processing and relating and structuring. The first two are related to stepwise processing while the last two map deep processing. Regulation strategies are metacognitive activities that students undertake. To map regulation strategies, the ILS-SV discerns three scales: external regulation, self-regulation and lack of regulation.

For all seven scales, the items are scored ranging from (1) 'I never or hardly ever do this' to (5) 'I (almost) always do this'. For each scale, Table 4.1 provides the number of items, an example item (translated from Dutch) and the estimate of reliability. Due to the sensitivity of Cronbach alpha to the number of items (Cortina, 1993; Palant, 2007), the mean inter-item correlation is a more appropriate measure of reliability for scales with few items (Briggs & Cheek, 1986). At each wave, all scales – each containing either 4 or 5 items – fall within the .2 to .5 range for good reliability (see Table 4.1).

Table 4.1: Learning strategy scales of the ILS-SV questionnaire, number of items, item examples (translated from Dutch) and range of scale reliability

Scales	Items	Item example	Mean inter-item correlation
Processing strategies			
Stepwise/surface			
Memorizing	4	I learn definitions by heart and as literally as possible.	.34-.39
Analysing	4	I study each course book chapter point by point and look into each piece separately.	.33-.36
Deep			
Critical processing	4	I try to understand the interpretations of experts in a critical way.	.32-.39
Relating and structuring	4	I compare conclusions from different teaching modules with each other.	.35-.46
Regulation strategies			
External regulation	5	I study according to the instructions given in the course material.	.20-.27
Self-regulation	4	I use other sources to complement study materials.	.28-.35
Lack of regulation	4	I confirm that I find it difficult to establish whether or not I have sufficiently mastered the course material.	.31-.38

4.2.3 Design, respondents and data availability

The research took place in a Flemish University College in which one cohort of students was followed. In March of the first academic year (from September to

June), all first-year students participated in the research during scheduled lecture slots. The same cohort was questioned again in May of the second and the third year. Each wave provided adequate response rates, as shown in Table 4.2. Of this cohort, 254 students participated in the three waves of data collection. 245 of those provided complete data on all ILS-SV items at each of the three waves. They constitute the longitudinal group for which growth is assessed in this study. The data are freely available upon request.

Table 4.2: Response rate per measurement wave

	Wave 1	Wave 2	Wave 3
Number of registered students	1412	731	561
Number of respondents	1047	515	392
Response rate (%)	74.1	70.4	65.8
Number of respondents with complete data	1037	507	363
Participants with complete data at each wave (longitudinal group)	245	245	245

As is common in longitudinal studies, not all students participated in the three waves of data collection. Firstly, some students stopped their studies or did not pass the exams, which is quite common in the first year of University College in Flanders (e.g. of the cohort under study, only 51.7% of the first year students were enrolled in the second year). For the first wave, independent-samples t-tests undertaken on the seven learning strategy scales (see Table 4.1), indicated that students from the longitudinal group (N=245) scored significantly higher on the analysing scale, but scored lower on the lack of regulation scale than their peers who participated in the first wave, but did not participate on all three occasions (N=802; $t(1045)=-3.217$, $p<.01$, Cohen's $d=.23$ and $t(1045)=5.55$, $p<.001$, Cohen's $d=.40$ respectively). For the second wave, comparable significant differences in the same two scales were noted (N=270, $t(513)=-2.7$, $p<.01$, Cohen's $d=.24$ and $t(510.39)=5.25$, $p<.001$, Cohen's $d=.46$ respectively). Although all effects are small (Cohen, 1988), the results suggest that students who persist in their studies through the third year do not constitute a random subset of students entering University College. This is in line with findings on the link between learning strategies and academic achievement (see Introduction). Therefore, for the analysing and lack of regulation scale, findings on the longitudinal group can only be generalised to the sub-population of students persisting in University College.

Secondly, some students persisted in their studies, but not in the research. Comparing students of the longitudinal group (N=245) to their peers answering at the third wave but not participating three times (N=124), reveals that the former score higher on the self-regulation strategy scale, though the effect is also small ($t(367)=-2.057$, $p<.05$, Cohen's $d=.23$). This result warrants caution in generalizing the findings for this scale to the subgroup of students persisting in University College.

4.2.4 MILG model[7]

To assess change in a learning strategy scale, we opted for a MILG model (Muthén & Muthén, 2010). Figure 4.1 exemplifies such a model, which consists of two levels.

---

[7] In chapter 2, a MILG was used as well (see 2.2.3). The difference is that here, longitudinal measurement invariance was tested for (see chapter 3), while this was not the case in chapter 2. For this reason, in Figure 4.1, the constraints on the factor loadings (λ's) and on the item difficulty level (thresholds, τ's) are included.

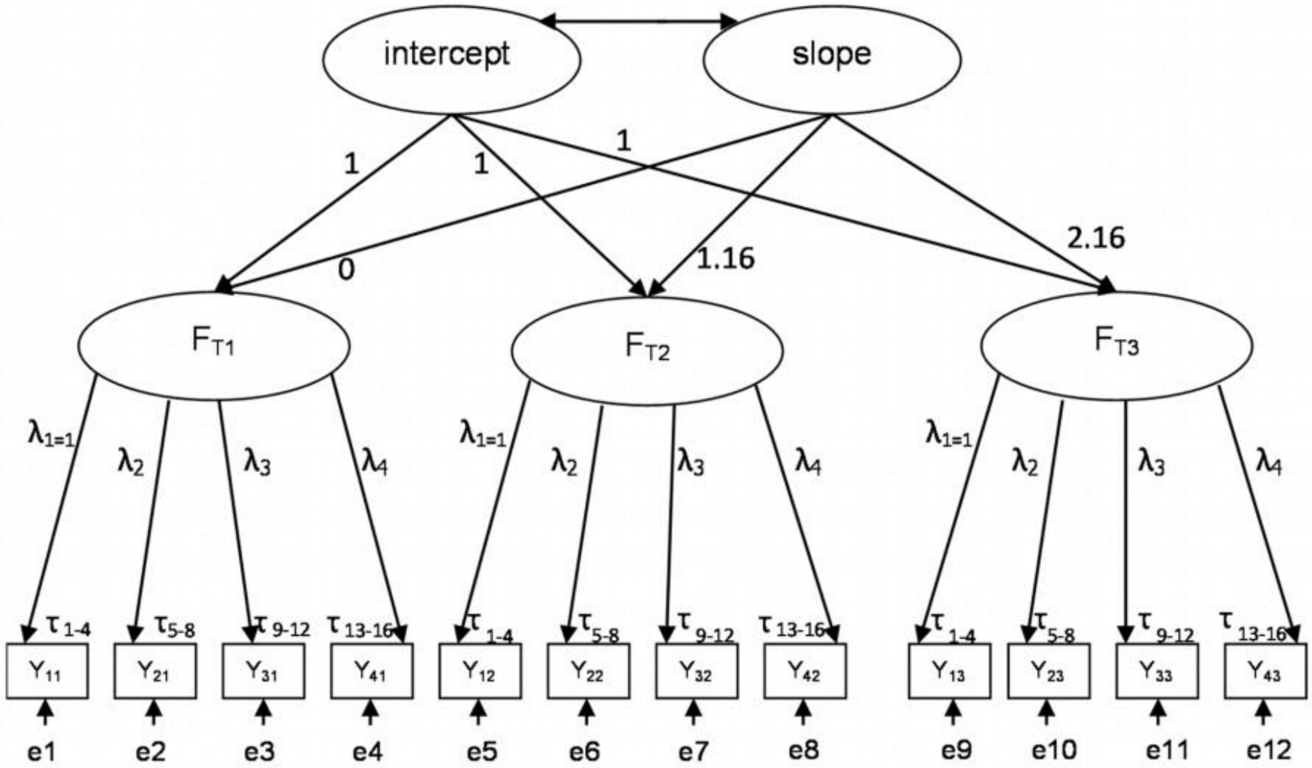


Figure 4.1: MILG model

The first level accounts for the common variation in the multiple indicators. A factor (e.g. the latent concept memorizing) is measured at three moments, using the same four items each time (Y_1-Y_4). An individual's score on an item at a certain point in time (Y_{ijt} e.g. Y_{i31} represents the score for individual i on the third item at the first wave) is predicted by a latent factor (e.g. F_{T1}). The second-order factors - intercept and slope - serve to explain the mean and covariance structure of these latent factors (Severiens et al., 2001). The intercept parameter signifies the average initial value for the scale (in our case: the value in March of the first year of University College), while the slope parameter estimates whether, on average, there is a significant increase or decrease in the scale scores per unit of time (in this case, 12 months). For more technical detail on latent growth modelling, we refer to Duncan et al. (2006) and Voelkle et al. (2006). Since the data consist of three waves, constant and linear models can be identified (Metha et al., 2004; Muthén & Muthén, 2010; Wu et al., 2010). Due to data gathering at unequal time intervals (14 months between wave 1 and 2 and 12 months between wave 2 and 3), the values of the factor loadings for the slope have been adjusted to 0, 1.16 (being 14/12th) and 2.16, respectively (Byrne, 2010; Muthén & Muthén, 2010).

Besides the estimates for average growth (intercept and slope), the latent growth analysis also details the differences between students (Duncan et al., 2006). A significant intercept variance indicates that students differ in their initial values on the scale, while slope variance suggests that students vary in their growth trajectory. If both variances prove significant, the covariance is informative as well (Byrne, 2010).

Next to this, the output of the MILG model details the explained variance in the latent factors. The R^2 is provided per latent factor, indicating which percentage of the variance in this latent factor is explained by the combination of the intercept and slope. Next to this, the MILG model output provides, for each latent factor, an estimate of the residual variance, indicating how much variance in this latent factor is left unexplained (Muthén & Muthén, 2010).

As mentioned above, making inferences about change in a learning strategy scale based upon repeated measurement, hinges upon the assumption of measurement invariance (Van der Veken et al., 2009; Volet et al., 1994; Zeegers, 2001). Longitudinal measurement invariance indicates that the definition of the latent construct is comparable over time (Stoel et al., 2006). In Figure 4.1, the two elements of this measurement invariance are depicted. Factor loadings are constrained equal over time (e.g. λ_2), indicating that one increase in a factor (F) represents the same increase in Y at different waves (Byrne, 2010). Next to this, the thresholds are presumed equal as well. With five-point Likert scales, there are four thresholds per item ($\tau_{1\text{-}4}$, see Figure 4.1). Threshold invariance implies that the percentage of students' choosing a Likert point should be comparable across waves (Metha et al., 2004).

Prior research has investigated this issue of longitudinal measurement invariance for the longitudinal sample under consideration (Coertjens et al., 2012). Results confirmed complete longitudinal measurement invariance for five learning strategy scales (memorizing, critical processing, relating and structuring, self-regulation and lack of regulation). With regard to the external regulation and the analysing scale, one and two thresholds, respectively, failed to reveal equivalence over measurement moments. These small measurement inequivalences are modelled in the partial measurement invariance models and taken into account when modelling growth for the external regulation and analysing scale.

MILG analyses were performed using Mplus 6.1. Due to the data's ordinal nature, the use of the conventional maximum likelihood (ML) estimation procedure could not be justified (Metha et al., 2004). The distribution free estimation procedure weighted least squares means-variance (WLSMV) was, therefore, employed (Fleming et al., 2008; Muthén & Muthén, 2010). Due to this WLSMV, the change in Chi^2 and degrees of freedom cannot be calculated in a straightforward fashion. "The difference in chi-square values for two nested models using the […] WLSMV chi-square values is not distributed as chi-square" (Muthén & Muthén, 2009b, p. 501). Therefore, a scaling correction (DIFFTEST function) is relied upon (Asparouhov & Muthén, 2006), of which only the p-value should be interpreted.

Though more simulation studies are required to establish the sample size requirements in the case of MILG models with ordinal data, research suggests the requirements do not differ compared to ML estimation. Moreover, the WLSMV was found to perform well with small samples, at least as well as the ML (Beauducel & Herzberg, 2006). Sample sizes of 200 or greater resulted in accurate estimates (Flora & Curran, 2004) on the condition that variables were not too skewed (Muthén, du Toit, & Spisic, 1997). The skewness of the ILS-SV items in this sample ranged between 0.012 and 1.134, and, following DiStefano and Hess (2005), indicated no reason for concern. Therefore, the sample of 245 students is adequate to estimate the MILG model.

In assessing the fit of each latent growth model, a series of fit indices was relied upon. However, as investigated by De Roche (2009), the cut-off values of these indices are not independent of the number of waves and respondents. In the case of three time points and 250 respondents, De Roche (2009) suggests examining the Chi², CFI, NNFI/TLI and RMSEA. However, the performance of the first may diminish with non-continuous data. Therefore, the CFI, NNFI/TLI and RMSEA are considered key. The first two proved robust in terms of sample size and number of waves, allowing the cut-off of .95 to be maintained. With regard to the latter, an adjusted cut-off of .08 is suggested (De Roche, 2009). Furthermore, we followed Wu et al.'s (2010) suggestion and examined the change in the CFI and

RMSEA. If, from the (complete or partial) measurement invariance model to the linear growth model, the CFI or the RMSEA deteriorates, this suggests that adding the growth factors did not help explain the patterns observed in the data (Wu et al., 2010).

4.3 Results

For each learning strategy scale, the fit of the MILG model is provided in Table 4.3. Table 4.4 shows the explained and residual variance at each wave for the seven scales whilst Table 4.5 presents the parameter estimates. The average growth trend for each scale is also displayed in Figure 4.2.

4.3.1 Processing strategies

For the memorizing scale, results indicate good fit for the linear growth model (Table 4.3). Moreover, compared to the invariant measurement model, fit did not deteriorate. At the different time points, the linear growth model succeeded in explaining between 65% and 68% of the variance in the latent factor. Nevertheless, as shown in Table 4.4, residual variance ranging between 18% and 20% remained at the three waves.

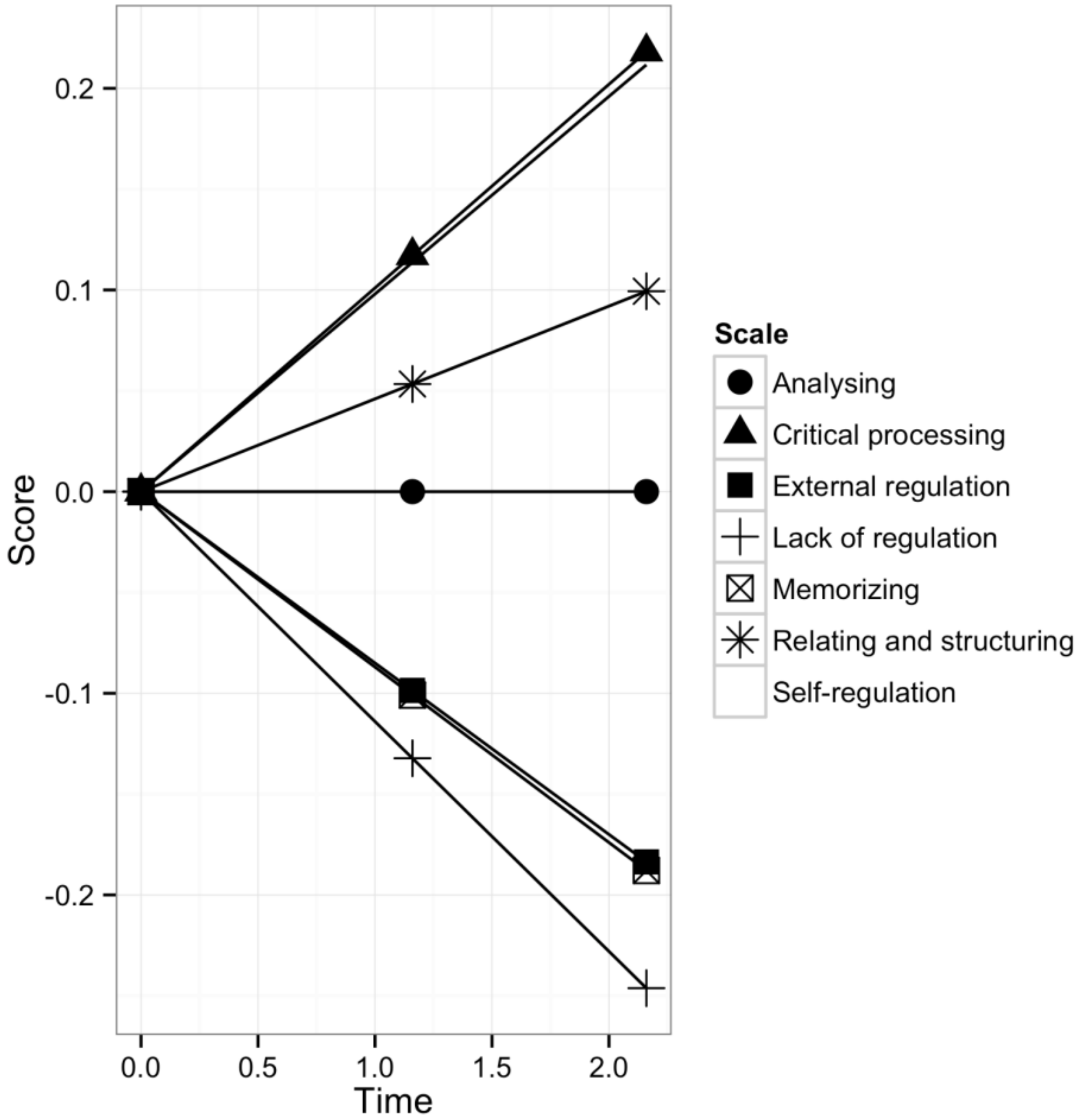


Figure 4.2: Average growth trajectories for the processing and regulation strategies

Table 4.3:Fit indices for measurement invariance and latent growth models

		χ^2	df	p	CFI	NNFI/TLI	RMSEA (90% conf. interval)
Memorizing	invariant measurement model	66.127	67		1.000	1.001°	.000-.037
	linear model	64.888	68		1.000	1.002°	.000-.034
Analysing	partial invariant measurement model	105.520	65	**	.969	.968	.032-.068
	linear model	106.209	66	***	.969	.969	.031-.067
Critical processing	invariant measurement model	82.600	67		.989	.989	.000-.051
	linear model	87.010	70		.988	.989	.000-.051
Relating and structuring	invariant measurement model	94.712	67	*	.985	.986	.019-.058
	linear model	109.679	70	*	.979	.980	.030-.065
External regulation	partial invariant measurement model	158.145	107	**	.953	.954	.029-.058
	linear model	159.781	110	**	.954	.956	.027-.057
Self-regulation	invariant measurement model	67.851	65		.998	.998	.000-.041
	linear model	66.983	66		.999	.999	.000-.039
Lack of regulation	invariant measurement model	103.456	67	*	.978	.979	.028-.064
	linear model	101.928	70	*	.981	.982	.023-.061

*** $p<.001$; ** $p<.01$; * $p<.05$; ° the NNFI/TLI can fall out of the 0-1 range when the df is larger than the χ^2 without implying an erroneous or just-identified model (De Roche, 2009)

Table 4.4: R^2 and residual variances (Res var) at the three waves

	Wave 1		Wave 2		Wave 3	
	R^2	Res var	R^2	Res var	R^2	Res var
Memorizing	.658	.198*	.652	.205***	.684	.189*
Analysing	.626	.191**	.793	.091*	.852	.076
Critical processing	.689	.128**	.679	.134**	.642	.159***
Relating and structuring	.671	.135**	.714	.11***	.67	.135***
External regulation	.593	.17***	.637	.142**	.598	.167***
Self-regulation	.506	.244**	.658	.193***	.929	.039
Lack of regulation	.669	.106***	.905	.022	.674	.104***

Note. All values of R^2 were significant at the .01 level; *** p<.001; ** p<.01; * p<.05; The R^2 and residual variance are not standardized and therefore do not add up to one.

Table 4.5: Parameter estimates for the MILG models

	Slope			VAR Intercept			VAR Slope		COV	
	Est.	SE	p	Est.	SE	p	Est.	SE	Est.	SE
Memorizing	-.087	.027	**	.381	.097	***	.010	.037	-.004	.051
Analysing	.003	.025		.319	.074	***	.027	.028	-.003	.036
Critical processing	.101	.022	***	.284	.047	***	put to zero		not estimated	
Relating and structuring	.046	.022	*	.274	.039	***	put to zero		not estimated	
External regulation	-.085	.024	***	.248	.041	***	put to zero		not estimated	
Self-regulation	.098	.028	***	.250	.087	**	.017	.034	.042	.042
Lack of regulation	-.114	.021	***	.215	.038	***	put to zero		not estimated	

*** $p<.001$; ** $p<.01$; * $p<.05$; Note that due to the MILG model, for each scale, the parameter estimate for the intercept is zero

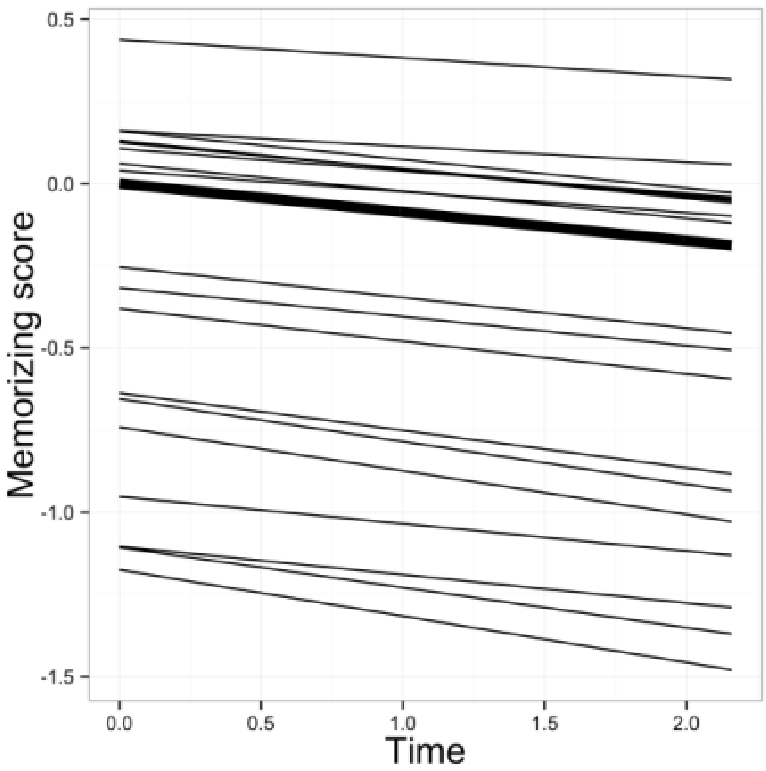


Figure 4.3: Average and predicted individual growth trajectories for the memorizing scale

The parameter estimates of the growth model are given in Table 4.5. For memorizing, a linearly decreasing reliance on the strategy during time in higher education is noted (*Est. slope*=-.087, *se*=.027, *p*<.01). This trend is depicted in Figure 4.2. The variance for the intercept proved significant (*Est VAR intercept*=.381, *se*=.097, *p*<.001, see Table 4.5), meaning that there are significant differences in students' initial levels of memorizing. Figure 4.3 shows the average growth trajectory as well as the predicted individual growth trajectory for a random subset of 20 students. It shows that there are differences at the start of higher education and that the general trajectory is not neatly followed by all students. The null hypothesis for the slope variance was, however, not rejected (*Est VAR slope*=.010, *se*=.037, *p*>.05).

As far as the analysing scale is concerned, indices suggested good model fit and remained at the same level compared to the partial invariant measurement model (see Table 4.3). The portion of the variance explained by the growth parameters increased from 62% to 85% over the waves, leaving significant residual variances only at the first two waves (see Table 4.4). As Figure 4.2 shows, the degree of analysing remains constant over time. Figure 4.4 shows for a random subset of 20 students the predicted individual growth trajectory in analysing next to average growth trajectory. The variance parameter for the slope did not prove significant, while students are estimated to vary in their initial level of analysing (see Table 4.5).

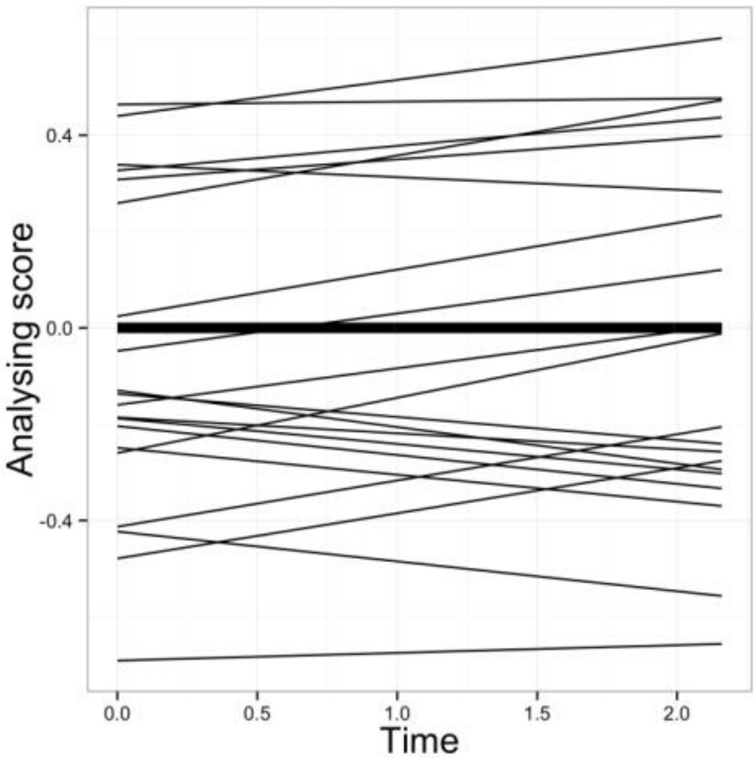


Figure 4.4: Average and predicted individual growth trajectories for the analysing scale

For the first of the deep processing scales, critical processing, results indicated a negative but non-significant variance for the slope (*Est. slope*=-.030, *se*=.027, *p*>.05). Following the suggestion by Muthén (2007), the variance for the slope was constrained to zero. Subsequently, the model showed good fit, also compared to the invariant measurement model (see Table 4.3). Parameter estimates suggest a linear increase in critical processing over time (see Figure 4.2). Concerning the differential growth, there was again significant intercept variance (see Table 4.5).

The results for the relating and structuring scale indicated a correlation between the intercept and the slope larger than one. To resolve this, the non-significant variance for the slope was constrained to zero. Successively, good model fit was found and the indices remained close to the level of the previous model (see Table 4.3). While significant residual variances were noted for the three latent factors, the model explains between 67% and 71% of the variance in latent factors scores (see Table 4.4). At the start of the study, students varied significantly in the degree to which they used relating and structuring. An increase in relating and structuring over time is noted as well (see Table 4.5 and Figure 4.2).

4.3.2 Regulation strategies

For the external regulation scale, the latent variable covariance matrix was not positive definite either, due to a negative, though non-significant, slope variance (*Est. VAR slope*=-.031, *se*=.028, *p*>.05). After putting this variance to zero, the fit of the latent growth model was adequate, and comparable to the partial invariant measurement model (see Table 4.3). The linear growth model explains between 59% and 63% of the variance in latent factor scores over time. Nevertheless, there is significant residual variance at each wave (see Table 4.4). Students are found to vary in their initial level of external regulation, and are noted to decrease their reliance on this regulation strategy over the course of their time in higher education (see Table 4.5 and Figure 4.2).

As far as the self-regulation scale is concerned, excellent fit was shown by the indices (see Table 4.3). Over the three waves, the R^2 improved from 50% to 93% (see Table 4.4). As shown in Figure 4.2, self-regulation increases over the course of this study. For 20 students, the individual predicted growth trajectory is shown in Figure 4.5. Examining the estimates for the variance parameters (see Table 4.5), a difference in students' initial levels of self-regulation is noted, while the general increasing trend can be assumed to be valid for all students.

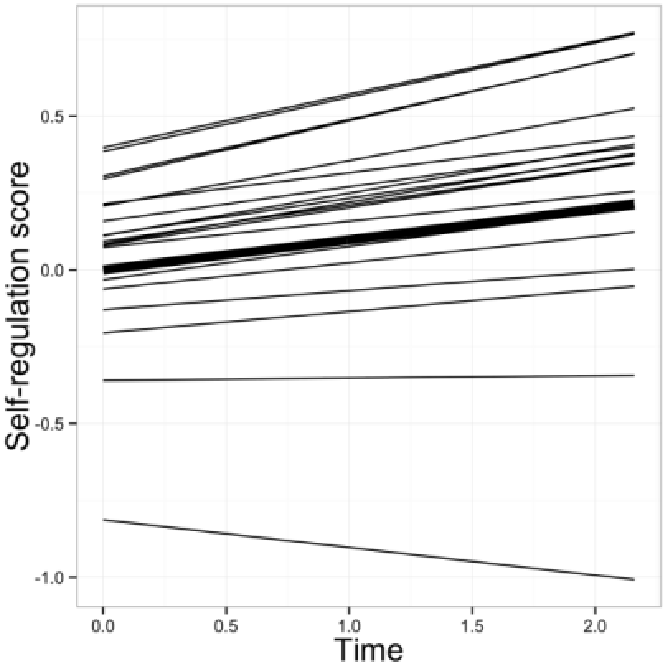


Figure 4.5: Average and predicted individual growth trajectories for the self-regulation scale

Lastly, the lack of regulation scale again showed a negative, though not significant, estimate for the slope variance (*Est. VAR slope*=-.012, *se*=.018, *p*>.05).

Constraining this variance to zero provided a good model fit, which was better compared to the fit for the invariant measurement model. Results suggest that students vary significantly in their initial score on this scale and decrease their lack of regulation during higher education (see Table 4.5 and Figure 4.2). The explained variance per latent factor ranges from 67% to 91%, leaving a significant residual variance of 10% at the first and at the third wave (see Table 4.4).

4.4 Discussion

Research on student learning has increasingly focussed on investigating the change in learning strategies during time spent in higher education. However, this domain relies predominantly on traditional statistical techniques, such as repeated measures ANOVA. The present research is innovative in the way that it investigates the average and differential growth trajectory in a more accurate and thorough fashion, through the use of MILG analysis.

The results regarding the average growth trajectories in processing and regulation strategies (RQ1) are in line with Donche et al.'s (2010) findings. Using the MILG model, results indicate that critical processing, relating and structuring, as well as self-regulation, increased over time. Memorizing, external regulation, and lack of regulation were found to decrease whilst the degree of analysing remained constant. However, for both the analysing and lack of regulation scales, the students persisting in their studies in a non-delayed manner form a biased follow-up sample. Therefore, the findings for these scales can only be generalized to the subgroup of students persisting in University College. For the self-regulation scale, significant differences were found between students participating in the third wave and those belonging to the longitudinal group. Taking this attrition into account, students following a non-delayed educational career in University College generally do seem to develop in the direction of more deep learning, moving away from surface-oriented and unregulated learning. Students persisting in the research as well seem to increase their self-regulation over time.

This finding of change in learning strategies over time adds to the trait-state debate. When taking measurement error, the ordinal nature of the Likert scale data and the partial measurement invariance adequately into account, results indicate variability in students' learning strategies over time. Thus, in line with prior longitudinal research (Busato et al., 1998; Gordon & Debus, 2002; Phan, 2011; Reid et al., 2005; Severiens et al., 2001; Vanthournout, 2011), the proposition of learning strategies as stable characteristics over time is refuted (Vermunt et al., 2014), even within an educational context that is relatively stable (i.e. the same University College).

Findings on differential growth (RQ2) indicate significant intercept variance for all scales, which is in line with prior research (Phan, 2011; Vanthournout, 2011). This finding suggests that in the longitudinal sample consisting of students persisting in their studies in a non-delayed manner and providing complete data at each wave (N=245), students vary in their initial degree of processing and regulation.

Next, the results suggest an absence of significant variability in slopes for all scales, contradicting prior preliminary analysis of the data, which relied upon discerning subgroups (Donche et al., 2010). The results are also at odds with studies in which advanced statistical analysis was used (Phan, 2011; Vanthournout, 2011). A first explanation could be a lack of statistical power required to reject the null hypothesis. Perhaps, there is differential growth, but given the sample size (N=245) significance could not be reached. A second explanation could be the selectivity of the longitudinal sample. Contrary to Vanthournout (2011) but in line with Phan (2011), only students providing complete data at each wave were retained in the present research. This choice stems from a limitation of MILG analysis, which does not allow for missing data at the item level.

To gauge the effect of attrition on the conclusions, the data were analysed in an alternative way. In contrast to the MILG analysis as presented, all students who participated at least once (N=1182) were included by relying on manifest scale scores and by modeling growth using a ML procedure. Results revealed that for four scales the conclusion of absence of slope variance remained (memorizing, relating and structuring, analysing and lack of regulation). For the critical processing scale and the external regulation scale, the slope variance was constrained to zero. For the self-regulation scale, however, the slope variance did just reach significance (*Est. VAR slope*=.047, *se*=.023, $p<.05$; *Est. covariance*=-.012, *se*=.030, $p>.05$). Thus, only for the self-regulation scale conclusions differed when considering attrition. Though not in line with prior findings (Phan, 2011; Vanthournout, 2011), for all other scales the evolution over time in learning strategies appears genuinely comparable. Further methodological research should focus on allowing missingness at the item level so that the strengths of the MILG can be combined with allowing for missing data. More research into the differential growth in learning strategies is warranted as well.

Concerning the explained and residual variance (RQ3), results indicate that between 50.6% and 92.9% of the variance in latent factors of processing and regulation strategy scales was accounted for by the linear growth trajectory. At the same time, for each of the learning strategy scales, at least two of the three errors variances proved significant. Though powerful predictors, the growth factors thus appear insufficient to predict the varying levels of the latent factors. On the one hand, these results plead for the use of nonlinear models in future research and, on the other hand, they warrant inclusion of other predictors to explain students' changes in learning strategies over time (Wu et al., 2010), such as prior education, gender or motivation to study.

Since the strength of any research study lies in the recognition of its limitations, two important constraints should be considered. The first is the issue of attrition, which is common to longitudinal studies, especially when longer time intervals are involved (Severiens et al., 2001; van der Veen & Peetsma, 2009). A second constraint concerns the number of measurement waves. With three data points, constant or linear growth trajectories can be estimated (Metha et al., 2004; Wu et al., 2010). However, more complex functions could clarify the issue of residual variances. Moreover, they are also of interest from a remedial point of view. Vermetten et al. (1999b), for example, suggested the need to investigate the developmental pattern in a more fine-grained fashion. 'Is it a one-way, gradual process in which students become more self-regulated, deep-level learners? Or is it a capricious pattern, with periods of stability followed by periods of change?' (p. 238) Thus, to better trace the actual development and detect opportunities for stimulating learning strategy development, future research should preferably span over longer periods of time (Mayer, 2011) and include more than three measurement points over time (Bijleveld, van der Kamp, & Mooijaart, 1998).

The constraints of the present study notwithstanding, results for the MILG analyses confirm the presence of a developmental trend in learning strategies during higher education towards high-quality learning. Students are, however, found to vary only in their initial levels of processing and regulation, but not in their development in these learning strategies during their time in higher education.

5. The growth trend in learning strategies during the transition from secondary to higher education

This chapter is based on Coertjens, L., Catrysse, L., Donche, V., De Maeyer, S., & Van Petegem, P. (2014, August). *(Predictors of) Growth in Learning Strategies during the Transition from Secondary to Higher Education*. Paper presented at the EARLI SIG 4 & 17 Conference, Leuven.

This study examines changes in learning strategies during the transition from secondary to higher education. It is hypothesized that students tend to move towards self-regulated and deep learning during this transition and that students' development over time varies from student to student for a limited number of learning strategy scales. All students from thirty-six secondary schools were logged onto the Inventory of Learning Styles-Short Version, and their progress was tracked over five waves from the beginning of the last year at secondary school to the beginning of their second year at a higher education establishment. Six hundred and thirty students were retained for analysis. Results indicate that students on average increased their self-regulated and deep learning during the transition. The results also showed an increase in students' degree of analysing and lack of regulation. Furthermore, for all the scales except the memorizing scale, the evolution over time varied from student to student.

5.1 Introduction

In recent years, participation in higher education has increased worldwide (Schofer & Meyer, 2005). However, this increased participation does not automatically result in increased successful completion of higher education studies. Though there are differences between countries, on average, one third of students entering higher education will not obtain a degree (OECD, 2013). It is clear that the first year of higher education is a major hurdle. Student dropout rate as well as non-success rate amongst students is highest during this first year (Bruinsma & Jansen, 2009; Hultberg, Plos, Hendry, & Kjellgren, 2008).

A number of researchers have explicitly devoted their attention to how students cope during the transition from secondary to higher education. One strand has examined the similarities of secondary and higher education (Torenbeek, Jansen, & Hofman, 2009; Torenbeek, Jansen, & Hofman, 2010). When the resemblance of the teaching/learning environment in secondary and higher education was high, students have declared to require less time to adjust, which affected achievement in a positive way. A second strand of research has focused on students' emotional experiences during the transition, describing a huge culture change and shock, accompanied by feelings of dislocation and stress (Christie et al., 2008). Moreover, when teaching approaches and learning environments are perceived to differ strongly from students' prior experiences, students have reported a loss of learning identity and have explained that they no longer feel competent as students. The authors therefore state that it is naive to assume that learning strategies from one school setting can be transferred to another. This transition effect is labeled 'learning shock' (Christie et al., 2008; Cree, Hounsell, Christie, McCune, & Tett, 2009).

Both strands concur that students' adjustment is impacted if the teaching/learning environment in higher education is perceived to be considerably different to the environment that they were used to. It has also been hypothesized that when students enter a new educational context, friction could incite students to adjust their way of going about learning to the new demands (Lindblom-Ylänne, 2003; Vermunt & Vermetten, 2004). It is to be expected, therefore, that the educational transition from secondary to higher education has an impact on students' learning strategies. However, research on the impact of the transition from secondary to higher education on students' learning strategies is currently sparse (Hultberg et al., 2008). This study attempts to answer a question that seems to remain unanswered by previous studies: is the 'learning shock' accompanied by a shock in learning strategies themselves? The present study will investigate whether and how students' learning strategies change during the transition from secondary to higher education. We will first discuss the learning strategies framework, and then present the research findings regarding growth in learning strategies during this transition. Finally, this study will formulate research hypotheses.

5.1.1 Learning strategies

The Students' Approaches to Learning (SAL) field provides a theoretical scope on how students learn (Coffield et al., 2004). Of all the questionnaires associated with SAL theories, Richardson (2004) distinguishes the two most frequently used with campus-based higher education students: the Study Process Questionnaire (SPQ; Biggs et al., 2001) and the Approaches to Study Inventory (ASI; Entwistle & Ramsden, 1983). From the list of SAL questionnaires, Vanthournout and colleagues (2014) discern a third that is also frequently used in higher education: the Vermunts' Inventory of Learning Styles Questionnaire (ILS). Using this framework, learning strategies are viewed as consisting of processing strategies (cognitive activities that a student applies whilst studying) and regulation strategies (i.e., metacognitive activities that students undertake, such as planning or testing oneself, Vermetten, Vermunt, & Lodewijks, 2002).

In this study we have opted for the last framework for three reasons. Firstly, the ILS framework takes a multidimensional approach to regulation strategies. Three scales are discerned: self-regulation, external regulation and lack of regulation. These strategies have repeatedly been found to be predictive of learning outcomes such as dropout and grade point average (Lindblom-Ylänne & Lonka, 1999; Vanthournout et al., 2012). In this light, examining how regulation strategies evolve during the transition from secondary to higher education is particularly relevant.

Secondly, compared to the SPQ and ASI frameworks, which discern a deep and a surface approach, the ILS framework offers a more detailed picture of students' processing strategies (Vanthournout et al., 2014). In a recent validation of the Inventory of Learning Styles – Short Version (ILS-SV, Donche & Van Petegem, 2008), four scales of cognitive processing strategies were distinguished: memorizing (1), analysing (2), critical processing (3) and relating and structuring (4). The first two are related to stepwise processing, while the last two to map deep processing.

Thirdly, the three frameworks concur that learning approaches or learning strategies are subjectable to change, but the time frame in which such growth can be expected differs. The SPQ and ASI framework predominantly focus on change in learning approaches during a course. The learning strategies in the ILS, mapping students' general preferences for learning which characterise a student for a certain period of time, are viewed as less context-specific (Vanthournout et al., 2014). Therefore, the ILS framework is more suitable for investigating growth trends in learning strategies for students in different teaching/learning environments, study domains and educational institutions.

5.1.2 Changes in learning strategies during the transition period

A number of studies have examined changes in learning strategies during higher education (see Vanthournout et al., 2011). However, research on the change in learning strategies during the transition period from secondary to higher education is currently lacking. There are a limited number of studies that elaborate on changes in learning strategies during a particular stage of the transition period, which is defined as the period of preparing for and adjusting to a new environment (Nicholson & West, 1995). The preparation phase takes place during the last year of secondary education, and the adjustment phase occurs during the first year of higher education (Torenbeek et al., 2010). Regarding the duration of the adjustment, experience with formal assessment at the higher education level was found to be crucial (Christie et al., 2008). Therefore, the transition period is considered to range from the last year of secondary education up to the start of the second year of higher education.

With regard to the last year of secondary education, no studies were found to use the ILS framework to map changes in learning strategies. With regard to the start of higher education up to the beginning of the second year, six studies use the ILS framework to map changes in learning strategies (Busato et al., 1998; Marambe, 2007; Severiens et al., 2001; Smith et al., 2007; Vanthournout, 2011; Vermunt & Minnaert, 2003). Regarding deep processing, Smith et al. (2007) and Marambe (2007) found a decreasing trend, while others detected a constant (Busato et al., 1998) or increasing trend (Severiens et al., 2001; Vermunt & Minnaert, 2003). Vanthournout (2011) concluded that only the degree of relating and structuring scale increases, while the degree of critical processing remains constant. Concerning stepwise processing, only Vermunt and Minnaert (2003) detected an increase while Busato et al. (1998) found a constant trend. Stepwise processing was mostly found to decrease over time (Marambe, 2007; Severiens et al., 2001; Smith et al., 2007). However, Vanthournout (2011) found that not all stepwise processing activities showed this trend. For instance, in this study, analysing remained constant while the reliance on memorising decreased over time.

Concerning the regulation strategies, self-regulation was generally found to increase (Severiens et al., 2001; Vanthournout, 2011; Vermunt & Minnaert, 2003), and in most studies, external regulation decreased over time (Severiens et al., 2001; Vanthournout, 2011; Vermunt & Minnaert, 2003). Lastly, results for the lack of regulation scale were mixed: lack of regulation was found to remain constant (Severiens et al., 2001) or to decrease over time (Vanthournout, 2011; Vermunt & Minnaert, 2003). For each regulation strategy scale, Marambe's study (2007) provided diverging results: a decreasing trend in self-regulation, a constant trend in external regulation and an increase in lack of regulation.

Thus, up to the start of the second year of higher education, the change in learning strategies tends to be a move towards the direction of self-regulated and deep learning. Although this appears to be at odds with research that suggests it is difficult to incite deep learning (Struyven, Dochy, Janssens, & Gielen, 2006; Vermetten et al., 2002), it is in line with prior findings on changes in learning strategies during higher education in general (Vanthournout et al., 2011).

5.1.3 *Differential growth in learning during the transition*

We turn our attention now to individual variations in student growth: can students be assumed to follow a comparable growth trajectory over time or not? Of the six studies described above, only one study has examined this differential growth (Vanthournout, 2011). For this reason, we broadened our scope to studies outside the transition period, resulting in one extra study (see chapter 4, Coertjens et al., 2013a) and to other SAL frameworks (SPQ; Phan, 2011), which provided a third study.

Phan's (2011) results indicated that students scoring lower on deep processing at the start of their undergraduate program increased their reliance on deep processing more rapidly. Vanthournout (2011) detected differential growth for the critical processing, self-regulation, analysing and external regulation scales. For the last two scales, this growth over time was related to students' initial score. Students scoring higher on analysing at the start of higher education tended to decrease their reliance upon it, while those initially scoring lower tended to increase their reliance upon it. For the external regulation scale, the findings suggested that students with a strong preference for external regulation at the start of their higher education decreased their reliance on external sources of regulation at a greater rate. Contrary to these findings, Coertjens et al. (2013a) did not detect differential growth for any scale during the three years of higher education (see chapter 4).

In sum, the limited studies on this topic tend to have varied conclusions with regard to differential growth in learning strategies. During higher education, students vary in growth over time for some scales and in some studies. For a subset of those scales, the differences between students decrease over time.

5.1.4 Research hypotheses

The present research aims to map students' average and differential growth in learning strategies during the transition from secondary to higher education. Given the lack of research on learning strategies during this transition, it is not suitable to generate classical hypothesis to be tested. The available research does, however, allow us to formulate some expectations in the form of a working hypothesis.

Previous qualitative and quantitative research has indicated that the transition from secondary to higher education is for most students an unsettling experience (Christie et al., 2008; Torenbeek et al., 2009). Students have to adapt to the demands of a new teaching/learning environment, which is expected to provoke a change in learning strategies. The limited past findings centred on the higher education part of the transition period and suggested an evolution towards deep processing and self-regulated learning. Therefore, we hypothesize that, *on average, students' learning strategies change during the transition from secondary to higher education in the direction of deep and self-regulated learning (Hypothesis 1).*

Limited past findings also indicate differential growth for a small number of learning strategies (Phan, 2011; Vanthournout, 2011), which, when related to the initial score, suggest students' results are more comparable over time. Therefore, we expect that *during the transition from secondary to higher education for a limited number of scales, students would evolve differently over time. If this change over time is* *related to students' initial score, it would indicate that students' scores become more comparable over time (Hypothesis 2).*

5.2 Method

5.2.1 Design and respondents

The data stems from a project on students' transition from secondary to higher education in Flanders (a Dutch speaking region in Belgium)[8]. All students in their final year of secondary education from thirty-six secondary schools offering a mixture of tracks (general, arts, technical and vocational) took part in the research project (N=3,704), which consisted of five waves as shown in Figure 5.1. During their last year of secondary education, students were questioned twice during school hours (wave 1: N=3,365; 91%, wave 2: N=2,839; 76,6%). At the second wave, students were also asked to fill out a consent form, and 84% complied. During an 18-month period after graduation, students were invited to participate three times (wave 3: N=1,101; 29.7%, wave 4: N=1,705; 46%, wave 5: N=1,029; 27.8%). At each of these waves, the participants received an email invitation to participate in the online questionnaire. As after two reminders via email the response rate was still low, the researchers called the respondents to ask them to complete the questionnaire up to three times. If the respondent did not respond the third time, a voicemail message was left.

---

[8] We acknowledge that the first three waves of this dataset have been previously reported on (see chapter 2, Coertjens et al., 2013b). However, in contrast to the present study, that article had a methodological focus. For this reason and given that in the present study the last two waves are included, we argue that the current study constitutes a new contribution to literature.

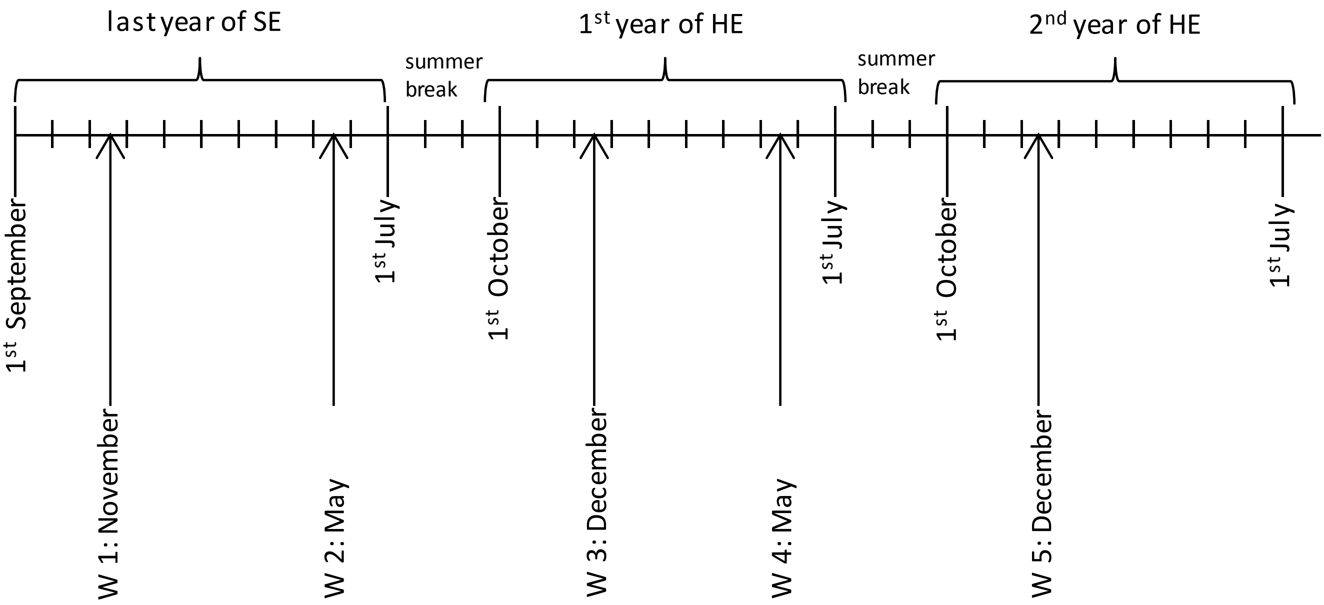


Figure 5.1: The five measurement waves over time

(SE=Secondary Education, HE=Higher Education)

In total 630 students declared themselves to be studying in higher education at waves three to five. Depending on the learning strategy, between 173 and 186 of the 630 students provided complete data. To give more detail on the amount of missing data, for the memorizing scale for example, 186 students provided complete data (29.5%). One hundred seventy eight (28.3%) had one missing data point, 174 (27.6%), 84 (13.3%) and 8 (1.3%) had respectively two, three and four missing data points. Methodological research on missing data in structural equation models, and latent growth models specifically, suggests not using listwise deletion. Including respondents with incomplete data by relying on a maximum likelihood estimation, for example, has been found to provide better results in terms of unbiased estimates and statistical power (Enders & Bandalos, 2001; Wothke, 2000). For this reason, the analyses were done on a sample of 630 students.

5.2.2 Measure

Students' learning strategies are investigated using the processing and regulation strategies of stemming from the ILS-SV, which has recently been validated for use on first-year Flemish University College students (Donche & Van Petegem, 2008). For all seven scales, the items are scored ranging from (1) 'I never or hardly ever do this,' to (5) 'I (almost) always do this'. For each scale, Table 5.1 provides the number of items, an example item and the range of scale reliability.

Table 5.1: Learning strategy scales of the ILS-SV questionnaire, number of items, item examples (translated from Dutch) and range of scale reliability

Scales	Items	Item example	Cronbach's Alpha
Processing strategies			
Stepwise processing			
Memorizing	4	I learn definitions by heart and as literally as possible.	.64-.74
Analysing	4	I study each course book chapter point by point and look into each piece separately.	.62-.69
Deep processing			
Critical processing	4	I try to understand the interpretations of experts in a critical way.	.69-.76
Relating and structuring	4	I compare conclusions from different teaching modules with each other.	.68-.72
Regulation strategies			
Self-regulation	4	I use other sources to complement study materials.	.61-.69
External regulation	6	I study according to the instructions given in the course material.	.56-.61
Lack of regulation	4	I confirm that I find it difficult to establish whether or not I have sufficiently mastered the course material.	.69-.75

Given that the learning strategy scales each have a small number of items (6 items for the external regulation scale and 4 items for other scales), which strongly affect the Cronbach's alpha (Cortina, 1993; Palant, 2007), a .60 cut-off is considered satisfactory. All scales show adequate reliability at each wave, except for the external regulation scale. Given that its reliability was below .60 at both the second and third wave, we refrain from modelling the evolution in external regulation during the transition from secondary to higher education.

5.2.3 Data analyses

In order not to confound true growth over time with change in the perception of the learning strategy questionnaire, longitudinal measurement invariance was tested for (Wu et al., 2010). The results are presented in Table 5.2 and confirm complete longitudinal measurement invariance for the analysing and critical processing scales. For the memorizing, relating and structuring and lack of regulation scales, one intercept failed to reveal equivalence over measurement moments. For the self-regulation scale, the constraints on two intercepts had to be freed. These small inequivalences are modelled in the partial measurement invariance models and were taken into account when modelling growth.

Table 5.3 shows the observed mean scores and standard deviations for the learning strategy scales. The growth patterns in the observed scores suggested possible non-linear growth over time for all of the scales. As suggested by Wang and Wang (2012) and Muthén and Muthén (2010) six latent growth models – one linear and five non-linear models - were estimated to adequately describe the growth trajectory for each scale.

Table 5.2: Results from measurement invariance tests for processing and regulation strategy scales

	Model description	χ^2	df	CFI	RMSEA	$\Delta\chi^2$	Δdf	p	ΔCFI
Memorizing	Baseline	193.195	120	.978	.031				
	Invariant loadings	205.149	132	.978	.030	11.954	12	.449	.000
	Invariant intercepts	251.972	144	.967	.034	46.823	12	***	-.011
	Partial invariant intercepts	232.647	143	.973	.032	27.498	11	**	-.005
Analysing	Baseline	203.253	120	.969	.033				
	Invariant loadings	213.372	132	.970	.031	10.119	12	.606	.001
	Invariant intercepts	242.740	144	.963	.033	29.368	12	**	-.007
Critical processing	Baseline	128.681	120	.997	.011				
	Invariant loadings	145.191	132	.996	.013	16.510	12	.169	-.001
	Invariant intercepts	177.784	144	.989	.019	32.593	12	**	-.007
Relating and structuring	Baseline model	142.889	120	.991	.017				
	Invariant loadings	157.471	132	.990	.018	14.582	12	.265	-.001
	Invariant intercepts	195.701	144	.980	.024	38.230	12	**	-.010
	Partial invariant intercepts	184.661	143	.984	.022	27.190	11	**	-.006
Self-regulation	Baseline	107.594	120	1.000	.000				
	Invariant loadings	143.992	132	.996	.012	36.398	12	***	-.004
	Invariant intercepts	231.525	144	.970	.031	87.533	12	***	-.026
	Partial invariant intercepts	208.897	143	.977	.027	64.905	11	***	-.019
	Partial invariant intercepts	158.247	142	.994	.013	14.255	10	.162	-.002
Lack of regulation	Baseline	231.953	120	.965	.038				
	Invariant loadings	239.098	132	.966	.036	7.145	12	.848	.001
	Invariant intercepts	284.547	144	.956	.039	45.449	12	***	-.010
	Partial invariant intercepts	264.506	143	.962	.037	25.408	11	**	-.004

*** $p<.001$; ** $p<.01$

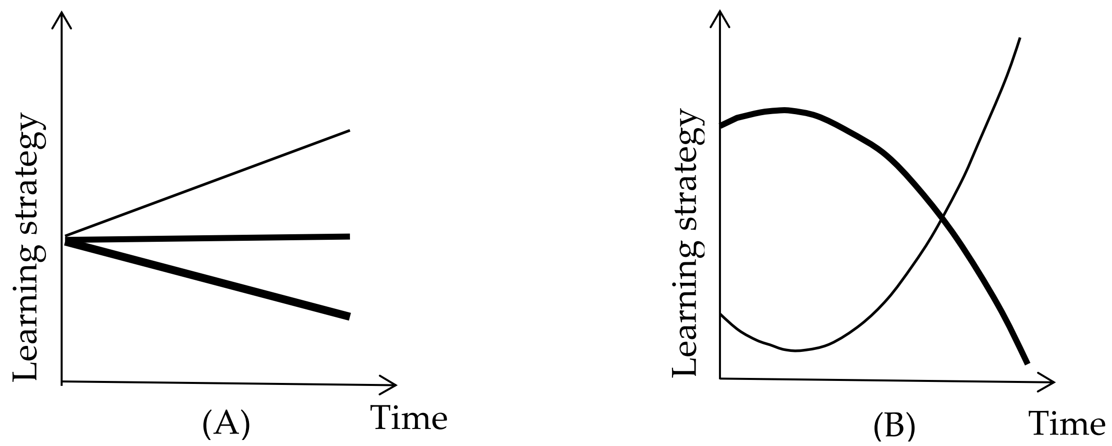
Table 5.3: Means and standard deviations for learning strategy scales

	November last year SE		May last year SE		December 1st year HE		May 1st year HE		December 2nd year HE	
	M	SD	M	SD	M	SD	M	SD	M	SD
Processing strategies										
Memorizing	3.47	0.78	3.34	0.79	3.52	0.78	3.45	0.78	3.41	0.80
Analysing	3.24	0.75	3.22	0.71	3.44	0.68	3.40	0.71	3.43	0.71
Critical processing	3.00	0.78	3.02	0.82	3.36	0.74	3.30	0.78	3.46	0.70
Relating and structuring	3.15	0.73	3.23	0.71	3.60	0.64	3.59	0.66	3.68	0.62
Regulation strategies										
Self-regulation	2.25	0.73	2.28	0.76	2.80	0.77	2.80	0.78	2.92	0.76
Lack of regulation	2.24	0.76	2.25	0.77	2.62	0.73	2.61	0.75	2.54	0.77

SE=Secondary Education, HE=Higher Education

First, a *linear* growth trajectory was estimated[9]. In such a model, the average growth over time is estimated using a straight line, defined by an intercept (i.e. the average score at the first wave) and a slope (i.e. the change in the scale per one unit of time). Examples of such increasing, decreasing or constant linear trends are given in Figure 5.2(a). Second, a *quadratic* growth trend is modelled. Next to the intercept and slope, a quadratic parameter is estimated, suggesting one bending point in the growth of a learning strategy over time (for examples of quadratic growth, see Figure 5.2(b)).

Third, a *cubic* growth trend (containing an intercept, slope, quadratic and cubic growth parameter) is assessed for how well it captures the growth in a learning strategy scale. With such a model, it is assumed that the growth in a learning strategy scale follows a trajectory with two bending points. For example, reliance on a learning strategy could initially decrease, then increase and by the end, decrease again (see Figure 5.2(c)).



[9] Due to data gathering at unequal time intervals (see Figure 5.1, respectively 6, 7, 5 and 7 months between the waves), the values of the factor loadings for the slope are adjusted to 0, 0.5, 1.08, 1.5 and 2.08 respectively (Byrne, 2010; Muthén & Muthén, 2010).

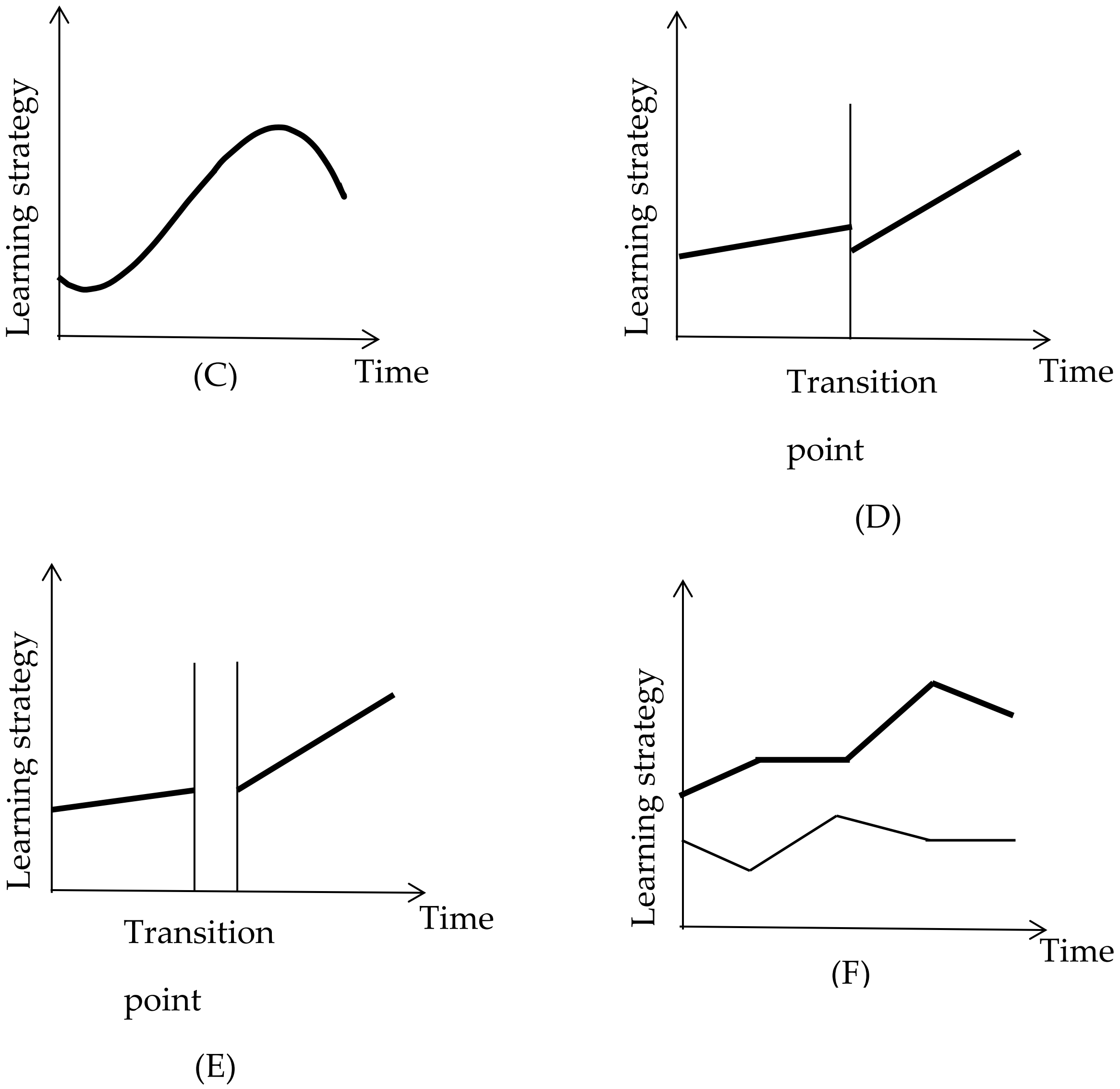


Figure 5.2: Example of the linear (A), quadratic (B), cubic (C), piecewise (D), discontinuous piecewise (E) latent growth model and a growth model with free time scores (F)

Fourth, a *piecewise growth model* is run. As shown in Figure 5.2(d), it approximates the nonlinear growth "through the use of two or more linear piecewise splines" (Bollen & Curran, 2006, p. 103). This is particularly useful for making comparisons in growth rates based on different developmental periods (Kim & Kim, 2012). In our study, two slopes are estimated (two-piece growth model), one for the two waves of secondary education and a second for the three waves of higher education. This reflects the expectation that the rate of change varies between the period prior to and after the transition.

Fifth, a *discontinuous piecewise growth model* estimates two different growth trajectories (Muthén & Muthén, 2010). As depicted in Figure 5.2(e), a separate intercept and slope are estimated for both secondary and higher education. As with the piecewise growth model, the slopes can be compared between the two educational periods. Next, it is possible to assess whether the leap due to the transition (i.e., the difference between the last score in secondary education and the first score in higher education) is significant (Kim & Kim, 2012).

The prior five models all assume that there is a certain predetermined trend that best captures growth in a learning strategy scale. The last model, a *growth model with free time scores*, relaxes this idea. Instead, the estimates model the capricious trend in the latent scale scores. This is done by freely estimating a number of time scores[10] (Bollen & Curran, 2006; Wu et al., 2010). Examples of such growth trends are given in Figure 5.2(f). Conceptually, this growth model discerns the time intervals in which the change in a learning strategy accelerates or decelerates (Muthén & Muthén, 2010; Wang & Wang, 2012; Wu et al., 2010). Next, it also allows the total change in a learning strategy to be partitioned over the time intervals. As such, the results suggest which time intervals are most important with regards to total change over time.

---

[10] In this study, the factor loadings of the slope for the first and the last time points are constrained to respectively 0 and 2.08 and the middle three factor loadings (or time points) are freely estimated ($\lambda 1$- $\lambda 3$).

The six models for each learning strategy scale were estimated using Mplus 6.1, using maximum likelihood estimation procedure. The model showing the lowest AIC and BIC was judged to best represent the observed trend (Grimm & Ram, 2009). Besides detailing average growth (see the six models above), the latent growth models also detail differential growth. First, the intercept variance indicates whether students vary with regard to their initial values on the scale. The slope variance suggests whether or not students differ how they change over time. If both variances prove significant, the covariance becomes a point of interest. A negative covariance suggests that students scoring lower on the intercept score higher on the slope (i.e., they display a stronger increase in the scale over time). This implies that students' results become more comparable over time. A positive covariance, on the other hand, suggests that students vary more towards the end the study than at the start.

5.3 Results

The fit of the latent growth models is described in Table 5.4 for each learning strategy scale. This table only displays the final accepted model for each scale, each of which provided good fit (CFI>.95; NNFI/TLI>.95; RMSEA≤.05; Byrne, 2010). Tables 5.5, 5.6 and 5.7 present the parameter estimates for the discontinuous piecewise growth model, the models with free time scores and the model with a cubic growth trend, respectively. Figure 5.3 visualises, for each scale, the growth trend as predicted by the latent growth model (white line). Next to this, to illustrate the variation in this growth trend, for a random subset of 150 students, the estimated growth over time is plotted as well.

5.3.1 Processing strategies

The change in memorizing scale is best captured by a discontinuous piecewise growth trend (Table 5.4). Prior to the transition from secondary to higher education, the average rate of change in the slope was -.313 (*se*=.060, *p*<.001), meaning that, on average, memorizing decreased .313 points over a 12 month period (Table 5.5). During the transition period, there was a significant leap in memorizing[11]. After the transition, the rate of change in the slope was -.104 (*se*=.032, *p*<.001), implying another, albeit less outspoken than during secondary education, decrease in memorizing, as visualised in Figure 5.3.

Regarding differential growth, the slope variance during higher education results non-significant (*est*=.063, *se*=.077, *p*>.05). This suggests that the general decreasing trend in memorizing during higher education can be assumed to hold for all students. We also note a positive association between the intercepts at secondary and at higher education (*est*=.267, *se*=.030, *p*<.001). This implies that students scoring higher on memorizing at the start of their last year of secondary education also score higher at the start of higher education. The covariances between the intercepts for secondary education and higher education on the one hand and the slope during higher education on the other hand were not significant (respectively, *est*=-.0.34, *se*=.022, *p*>.05 and *est*=.029, *se*=.036, *p*>.05), indicating that students' growth during higher education is not systematically related to their scores at the wave 1 (start of the last year of secondary education)

[11] Whether this leap is significant or not, was tested by re-arranging the factor loadings of the slope in SE in a manner that the intercept was at the end of SE (second wave). In this way, the model compared the intercept in HE (3rd wave) to the value at the second wave. This resulted significant (*est*=.146, *se*=.031, *p*<.001).

or wave 3 (start of higher education). This implies that the variation between students in terms of their degree of memorizing remains constant over time.

For the analysing scale, a growth model with free time scores best fitted the data (see Table 5.4). The parameter estimates for this model are provided in Table 5.6 and the growth trend is visualised in Figure 5.3. The results indicate an increasing trend in the analysing scale (*est*=.088, *se*=.014, *p*<.001). Looking at when this growth occurs, results show a constant trend in analysing secondary education (from wave 1 to 2; λ_1, *est*=-.216, *se*=.277, *p*>.05). To discern whether there is a significant change during the transition period, the difference between λ_2 and λ_1 is divided by the standard deviation in λ_2. If the result exceeds 1.96, the lambda is significantly different from the previous (Muthén & Muthén, 2010), indicating a significant change during the time form wave 2 to 3. Here, the degree of analysing is found to increase during the transition ((2.033-0/0.256)>1.96). Afterwards, between December and May of the first year of higher education (from wave 3 to 4), and from May of the first year of higher education to December of the second year (from wave 4 to 5), the degree of analysing anew remains constant (calculated as (1.869-2.033)/0.238>1.96 and (2.08-1.896)/0.238>1.96, see Table 5.6, Muthén & Muthén, 2010). Thus, the results suggest that there is an increase in the degree of analysing, but that this increase occurs during the transition from secondary to higher education.

Table 5.4: Fit indices for the best fitting latent growth model per learning strategy

	Accepted model	χ^2	df	p	CFI	NNFI/TLI	RMSEA (90% conf. interval)
Memorizing	Discontinuous piecewise growth model	234.815	148	***	.973	.966	.031 (.023-.038)
Analysing	Growth model with free time scores	244.005	151	***	.965	.956	.031 (.024-.038)
Critical processing	Growth model with free time scores	194.020	151	*	.986	.983	.021 (.011-.030)
Relating and structuring	Growth model with free time scores	190.965	150	*	.984	.980	.021 (.010-.029)
Self-regulation	Growth model with free time scores	167.097	149	.148	.994	.992	.014 (.000-.024)
Lack of regulation	Cubic growth trend	296.002	149	***	.954	.941	.040 (.033-.046)

*** $p<.001$; * $p<.05$

Table 5.5: Parameter estimates for the discontinuous piecewise growth model

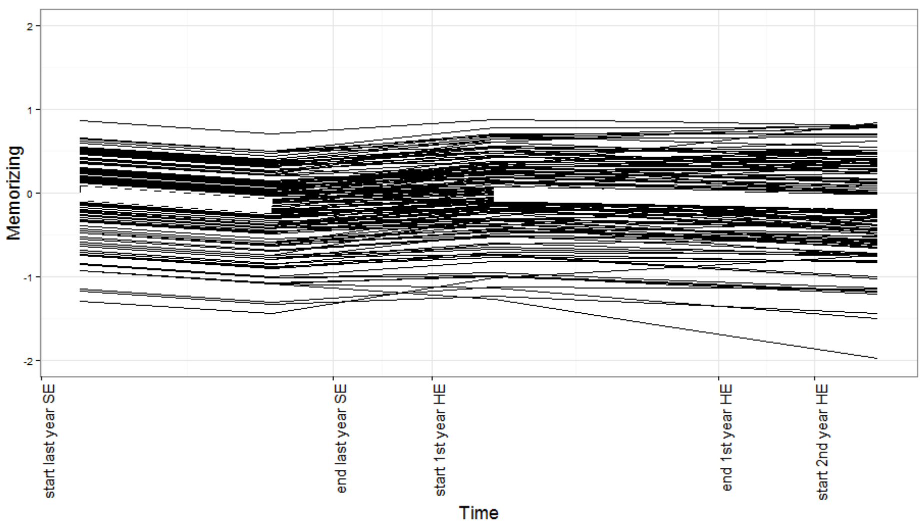
	Slope SE	Intercept HE	Slope HE	Var Intercept SE	Var Intercept HE	Var slope HE	Cov Intercepts SE & HE	Cov Intercept SE & slope HE	Cov Intercept HE & slope HE
Memorizing	-0.313 (.060)***	-0.010 (.031)***	-0.104 (.032)***	0.278 (.032)***	0.281 (.043)***	0.063 (.077)	0.267 (.030)***	-0.034 (.022)	0.028 (.036)

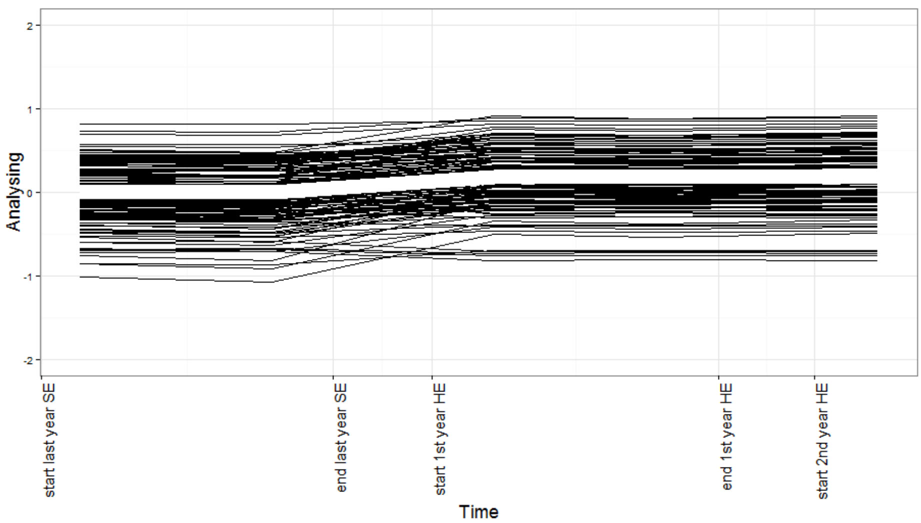
SE=Secondary Education, HE=Higher Education; *** p<.001; ** p<.01; Note1: In a latent growth model with underlying measurement model (to test for measurement invariance over time), the intercept during SE is zero. Therefore, it is not provided. Note2: the slope variance during SE is constrained to zero, given there were two measurement moments during SE.

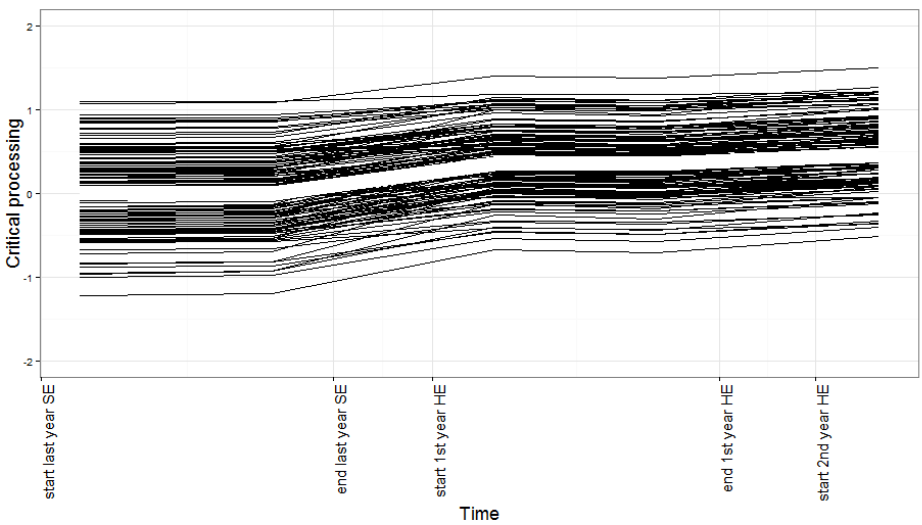
Table 5.6: Parameter estimates for the growth models with free time scores

	$\lambda 1^{\circ}$	$\lambda 2^{\circ}$	$\lambda 3^{\circ}$	$\lambda 4^{\circ}$	Slope	Var Intercept	Var slope	Cov
Analysing	-0.216 (.277)	2.033 (.256)*	1.869 (.238)	2.08 (.000)	0.088 (.014)***	0.190 (.027)***	0.024 (.006)***	-0.024 (.009)**
Critical processing	0.060 (.131)	1.609 (.122)*	1.491 (.121)	2.08 (.000)*	0.221 (.018)***	0.371 (.040)***	0.038 (.009)***	-0.050 (.014)***
Relating and structuring	0.393 (.100)***	1.767 (.088)*	1.816 (.088)	2.08 (.000)*	0.284 (.020)***	0.329 (.040)***	0.061 (.011)***	-0.070 (.017)***
Self-regulation	0.085 (.097)	1.619 (.114)*	1.705 (.107)	2.08 (.000)*	0.302 (.020)***	0.274 (.032)***	0.051 (.011)***	-0.007 (.014)

*** $p<.001$; ** $p<.01$; * $p<.05$; ° Mplus provides a significance test for the λ's, indicating whether to the λ is significantly different from zero. This is relevant for $\lambda 1$, given that the factor loading for wave 1 is zero. For the other λ's, it is of more interest whether the score differs from the previous λ, suggesting whether there is an increase or decrease in the scale. This is calculated by the difference in λ/the standard deviation. If the result exceeds 1.96, the lambda is significantly different from the previous (Muthén & Muthén, 2010). For example, for $\lambda 2$ of the critical processing scale: (1.609-0.060)/0.122= 12.7 > 1.96. For the growth from wave 4 to 5 ($\lambda 3$ to $\lambda 4$), we use the standard error of $\lambda 3$.
In table 5.6 we provide this significance, instead of the results provided in the Mplus output;
Note: In a latent growth model with underlying measurement model (to test for measurement invariance over time), the intercept is zero. Therefore, it is not provided.


Memorizing
2
1
0
-1
-2
start last year SE
end last year SE
start 1st year HE
end 1st year HE
start 2nd year HE
Time


Analysing
2
1
0
-1
-2
start last year SE
end last year SE
start 1st year HE
end 1st year HE
start 2nd year HE
Time


Critical processing
2
1
0
-1
-2
start last year SE
end last year SE
start 1st year HE
end 1st year HE
start 2nd year HE
Time

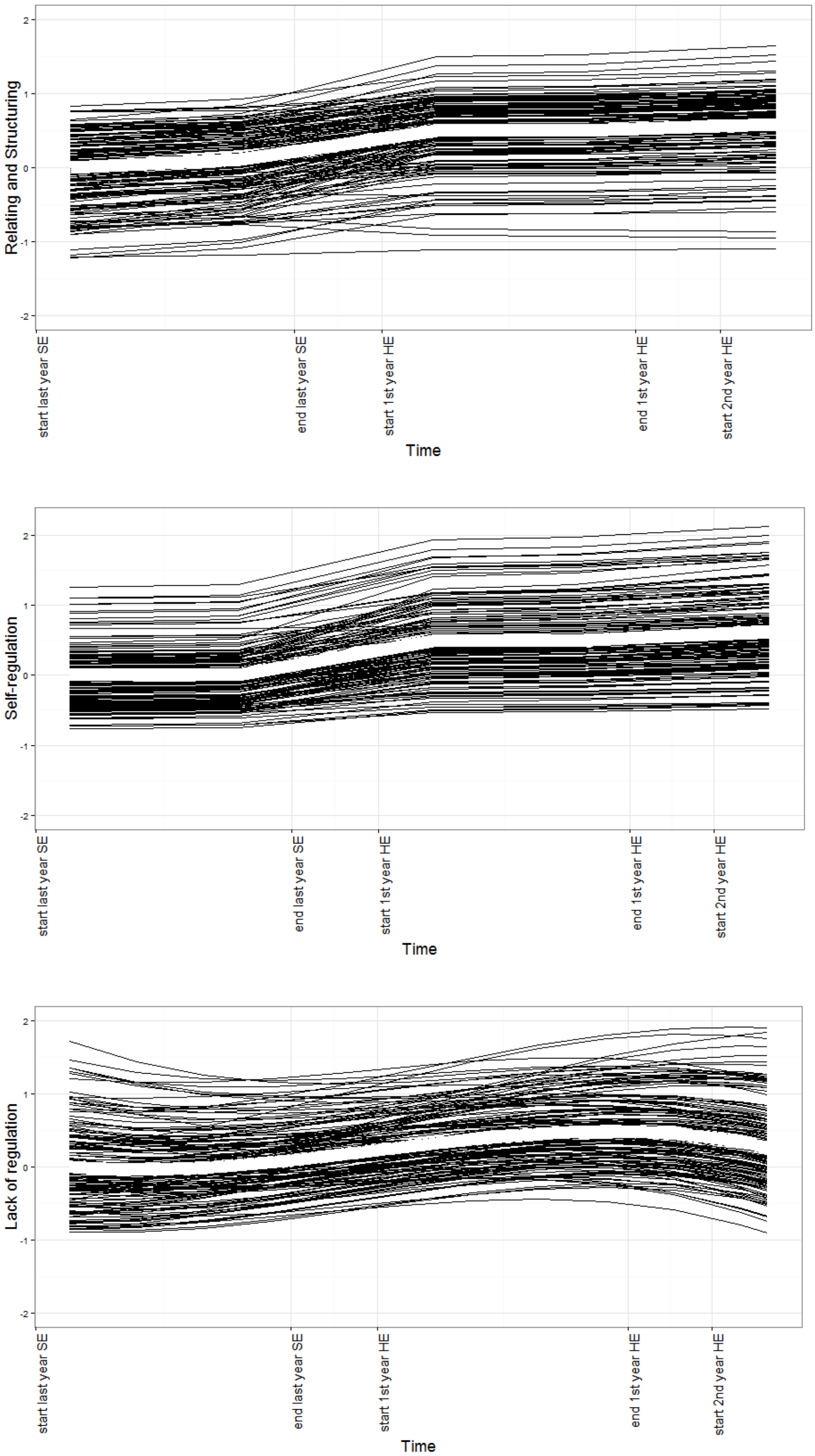


Figure 5.3: Average and individual estimated growth trajectories per learning strategy scale

Regarding the differential growth over time in the analysing scale, results indicate a significant slope variance (*est*=.024, *se*=.006, *p*<.001). This implies that students vary in their growth in analysing over time. Next, the covariance between the intercept and slope is significant and negative (*est*=-.024, *se*=.009, *p*<.01), suggesting that students' scoring higher on analysing at wave 1 increase their reliance on this processing strategy, but to a lesser degree. The same applies in the other direction: students' scoring lower at wave 1 show a greater increase in their degree of analysing. Thus, students' scores on the analysing scale become more comparable over time.

For the critical processing scale, a constant trend is found during the students' last year of secondary education (from wave 1 to 2; λ_1, *est*=.060, *se*=.131, *p*>.05, see Table 5.6). During the transition, there is a significant increase in critical processing ((1.609-0)/0.122>1.96). The remainder of the first year of higher education, critical processing remains constant again (1.491-1.609)/0.121<1.96). From the end of their first year of higher education to December of their second year of higher education critical processing increases anew ((2.08-1.494)/0.121>1.96), though, as Figure 5.3 shows, less strong than during the transition. In sum, the growth in critical processing occurs during the transition phase and from the end of the first year of higher education to December of the second year of higher education.

Examining the parameter estimates on differential growth reveals that students vary in their growth over time (*est*=.038, *se*=.009, *p*<.001). Moreover, there is a negative covariance between the intercept and slope (*est*=-.050, *se*=.014, *p*<.001), implying that, students' scores on critical processing become more alike over time.

Regarding the relating and structuring scale, results indicated the best fit for a latent growth model with free time scores (see Table 5.4). Overall, there is an increase in relating and structuring over time (*est*=.284, *se*=.020, *p*<.001, see Table 5.6). Looking into detail on when this growth occurs, it is noted that during the last year of secondary education, relating and structuring increases (*est*=.393, *se*=.100, *p*<.001). During the transition period relating and structuring augments as well ((1.767-0.393)/0.088>1.96). As shown in Figure 5.3 this increase is at a faster rate than during the last year of secondary education. During the remainder of the first year of higher education, the degree in relating and structuring remains constant ((1.816-1.767)/0.088<1.96). From the end of the students' first year of higher education to December of their second year, there is a slight increase in relating and structuring ((2.08-1.816)/0.088>1.96).

Regarding differential growth, the change in relating and structuring over time was found to differ between students (*est*=.061, *se*=.011, *p*<.001). The covariance resulted significant as well (*est*=-.070, *se*=.017, *p*<.001), indicating that students scoring higher on relating and structuring at the first wave (November of the last year of secondary education) tended to increase their relating and structuring to a lesser degree and vice versa.

5.3.2 Regulation strategies

For the self-regulation scale, a growth model with free time scores was most suitable (see Table 5.4). The parameter estimates indicate that on average over a period of 12 months self-regulation increases by .302 (*se*=.020, *p*<.001, see Table 5.6). Examining when this growth occurs, reveals that during secondary education, self-regulation remains constant (*est*=.085, *se*=.096, *p*>.05). As shown in Figure 5.3, during the transition from secondary education to higher education, there is a significant jump in self-regulation ((1.619-0)/0.114>1.96). During the remainder of students' first year of higher education, self-regulation remained constant ((1.705-1.619)/0.107<1.96), to subsequently increase anew from the end of the students' first year of higher education to the start of their second year of higher education ((2.08-1.705)/0.107>1.96), though less strongly then during the transition (see Figure 5.3).

Concerning differential growth in self-regulation, students were found to differ in their growth over time (*est*=.051, *se*=.011, *p*<.001, see Table 5.6). The insignificant covariance (*est*=-.007, *se*=.014, *p*>.05) indicates that this growth is unrelated to the students' score on self-regulation in November of their last year of secondary education.

Table 5.7: Parameter estimates for the cubic growth model

	Linear	Quadratic	Cubic	Var Intercept	Var Linear	Var Quadratic	Cov Intercept & Linear	Cov Intercept & Quadratic	Cov Linear & Quadratic
Lack of regulation	-0.384 (.121)**	1.034 (.153)***	-0.379 (.049)***	0.453 (.065)***	0.406 (.175)*	0.061 (.028)*	-0.243 (.096)*	-0.060 (.036)	-0.131 (.067)

*** $p<.001$; ** $p<.01$; * $p<.05$; Note: In a latent growth model with underlying measurement model (to test for measurement invariance over time), the intercept is zero. Therefore, it is not provided.

Last, for the lack of regulation scale, a cubic growth trend represented the data best (see Table 5.4). The results of this model indicate a negative linear slope (*est*=-.384, *se*=.121, *p*<.01, see Table 5.7), a positive quadratic parameter (*est*=1.034, *se*=.153, *p*<.001) and a negative cubic growth parameter (*est*=-.379, *se*=.049, *p*<.001). To interpret these, the visualisation of growth trend in Figure 5.3 can be helpful. From this, it appears that during the last year of secondary education, students' lack or regulation remains constant. During the transition period up to the start of the second year of higher education, the degree of lack of regulation increases. Subsequently, up to December of the second year of higher education, students' lack of regulation decreases.

Regarding differential growth in the lack of regulation scale, results indicate that students vary in their linear and quadratic growth (respectively *est*=.406, *se*=.175, *p*<.05 and *est*=.061, *se*=.028, *p*<.05, see Table 5.7). The variance of the cubic growth trend was insignificant and had to be constrained to zero to improve the model fit. Results also show that students scoring higher on the lack of regulation scale at the start of the last year of secondary education show a stronger decrease during the last year in secondary education (*est*=-.243, *se*=.096, *p*<.051). Thus, as shown in Figure 5.3, over the last year of secondary education, the differences between students in terms of lack of regulation diminished.

The covariance between the intercept and the quadratic growth parameter was not significant (*est*=-.060, *se*=.036, *p*>.05, see Table 5.7), while the covariance between the linear slope and the quadratic growth parameter was at the verge of significance (*est*=-.131, *se*=.067, *p*=.053). Interpreting this as significant, this suggests that students showing a less steep decreasing linear slope tended to score lower on the quadratic growth parameter. In other words, students who decreased less in their lack of regulation during their last year of secondary education tended to show a more modest increase in their lack of regulation during the transition and their first year of higher education. The same applies in reverse: students who initially decreased stronger on the scale later showed a stronger increase. Therefore, and in contrast to the last year of secondary education, during the transition and their first year of higher education, students' scores on lack of regulation become less comparable over time.

5.4 Discussion

Research on the experience of the transition period from secondary to higher education highlights the importance of the change in students' teaching/learning environment (Christie et al., 2008; Torenbeek et al., 2010). This may affect students' learning strategies. The present research is innovative in the way that it investigates the average and differential growth in learning strategies during the transition from secondary to higher education, by relying on longitudinal data from the onset of the students' last year of secondary education up to halfway through their second year in higher education.

5.4.1 Results on average growth

Results partially confirm hypothesis 1 regarding an average change in the direction of deep and self-regulated learning. There is an increase in critical processing, relating and structuring as well as self-regulation during the 25 months of the study. This increase is more pronounced during the actual transition from secondary to higher education. With regard to the memorizing scale, students displayed a sharp increase in their degree of memorizing during the transition. Yet, due to the decreasing trends during both secondary and higher education, there was an overall decrease from the first to the last wave (see Figure 5.3). At the start of their second year of higher education, students, on average, relied less on memorizing than at the start of their last year of secondary education. The increase in deep and self-regulated learning and the decrease in memorizing fit with the hypothesis.

The results for the analysing and lack of regulation scales, however, contradict the hypothesis. Analysing was found to increase during the transition. Students' lack of regulation increased during the transition from secondary to higher education as well as during their first year of higher education. Although it later decreased, the student's lack of regulation at the last wave was higher than at the first wave. In sum, the results partially confirm hypothesis 1: there is an increase in deep and self-regulated learning as well as in analysing and unregulated learning, and a decrease in memorizing.

5.4.2 Results on differential growth

The second hypothesis stated that differential growth would be detected for a limited number of scales (1), which, if related to the initial level, would indicate that students' scores become more comparable over time (2). With regard to part (1), this hypothesis proved to be too modest: differential growth over time was found for all scales except the memorizing scale. This implies that students vary in how they change these strategies from their last year of secondary education to halfway through their second year of higher education.

The fact that the hypothesis proved to be too modest could well be due to the power of the model, which is related to the number of waves (Wu et al., 2010). In the study by Coertjens et al. (2013a), which included three waves, no differential growth was detected. Phan (2011) and Vanthournout (2011), relying on four measurement points, found differential growth in some scales. Here, with five waves, differential growth in all except one scale is detected. Future research could attempt to explain the slope variance by including predictors for the slope.

Results regarding part (2) of hypothesis 2 suggest that, except for the self-regulation scale, the change over time is related to the score on the intercept. For the analysing, critical processing and relating and structuring scales, students' scores became more comparable over time. Students' scores for the lack of regulation scale also became more similar during the last year of secondary education. However, during the transition and during their first year of higher education, differences between students in terms of their lack of regulation increased again. In sum, only the results of the differential growth for the lack of regulation scale contradict the second hypothesis.

5.4.3 Limitations and future research

A limitation of the present study concerns the attrition issue common to longitudinal studies, especially when long time intervals are involved. Some students were unreachable during the three waves after secondary education.

For them, it was unclear whether they took on a job, did something else (e.g., went travelling) or were studying in higher education. Therefore, these students were not included in the study. This attrition issue and the fact that the present study is the first to examine the change in learning strategies during the transition period from secondary to higher education, underlines this need for replication research to confirm the findings.

The present research set out to map students' growth in learning strategies during the transition from secondary to higher education. To better understand why the observed change in learning strategies occurs, in future research, this change in learning strategies could be related to other concepts. Here, students' perceptions of the teaching-learning environment appear a particularly valuable one. There is an extensive research base of cross-sectional studies, relating both concepts at a certain point in time (e.g., Kreber, 2003; Parpala, Lindblom-Ylänne, Komulainen, Litmanen, & Hirsto, 2010). It would be of interest to re-assess this relationship in a longitudinal design: is the change in students' perception of the teaching-learning environment related to the change in their learning strategies?

Notwithstanding this constraint, the present study suggests that, during the transition from secondary to higher education, students, on average, increase their deep and self-regulated learning to the detriment of memorizing. However, the degree of analysing as well as unregulated learning also increased. Concerning differential growth, results indicate that for all but the memorizing scale, students differed in their growth over time. For the analysing, critical processing and relating and structuring scales, students' scores became more comparable over time. However, differences between students in terms of lack of regulation increased during the transition to and during their first year of higher education. Concluding, the 'learning shock' as described by Christie et al. (2008) appears to be accompanied by average as well as differential growth in learning strategies.

6. Sensitivity of the growth trend estimates to the missing data technique

This chapter is based on Coertjens, L., Donche, V., De Maeyer, S., & Van Petegem, P. (2012, August). *Is missingness in longitudinal research during higher education dependent on student's learning strategies? A Flemish case study*. Paper presented at EARLI SIG 1 conference, Brussels

Longitudinal data is fraught, almost always, with missing data. However, in educational research, there is a large discrepancy between the methodological suggestions and research practice. Whilst the former suggests applying sensitivity analysis to assess the robustness of the results to varying assumptions regarding the mechanism generating the missing data, usually in practice, missing data is ignored by relying on listwise deletion. There are few practical examples of sensitivity analysis in the educational research domain. Taking the case of the changes in learning strategies during higher education, this study provides a tutorial example of sensitivity analysis. In a Belgian university college, one cohort of students was asked to complete the Inventory of Learning Styles – Short Version in three measurement waves. A substantial number of students did not participate on each occasion due to non-response or attrition. Therefore, the analysis consisted of two studies: one on students continuing in higher education (with different degrees of non-response), and a second on all available data showing missingness due to attrition and non-response. The results indicated that, for some learning strategy scales, growth estimates differed between models assuming different mechanisms for missingness. Moreover, those varying results suggested substantively different conclusions. Guidelines are provided in reporting the results from sensitivity analysis.

6.1 Introduction

In the educational research domain, longitudinal designs are relied upon to assess, for example, how achievement goals evolve during the transition from elementary to secondary school, how reading comprehension evolves after an intervention, how the relation between perceived mastery goal structures and perceived teacher support changes across the school year, or how student learning alters during higher education (Zeegers, 2001). The types of data, gathered in such longitudinal designs, invariably contain a certain amount of mig data: not all respondents participated in every wave. Some respondents dropped out of the study before completion (attrition), whilst others may miss a wave but return for a subsequent one (wave non-response). A third group may participate in a data collection but leave a number of questions or items unanswered (item non-response).

This missing data is a major issue in educational and psychological research. Peugh and Enders (2004) found that, in longitudinal studies in these research domains, on average, 9.78% of the data was missing and this could grow to a maximum of 67%. When examining studies on the growth in student learning strategies during higher education, this percentage was even higher: between 43% (Van der Veken et al., 2009) and 93.5% (Jackling, 2005) of students, participating in the first wave, did not participate in all waves.

Moreover, as acknowledged in numerous longitudinal studies, the missing data could generate a biased follow-up sample. Such selective samples can be a threat to internal validity (Foster et al., 2004; Shadish et al., 2002). Perhaps different conclusions would have been reached had the percentage of missing data been (a lot) smaller.

In the light of internal validity, the mechanism, generating the missing data, is found to be crucial. Students can be missing due to chance (e.g., having the flu at the time of the test, Schünemann, Spörer, & Brunstein, 2013) but this can also be related to the concept under study. For example, pupils scoring higher on performance approach goals may be more likely to stay back a grade in secondary education; students with low reading comprehension may be less willing to participate in research on the topic; and students, with a less desirable learning strategy, have a higher chance of dropping out of higher education. When missingness is indeed related to the concept under study, internal validity may be threatened (Foster et al., 2004).

In literature, three mechanisms generating missing data are discerned: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) (Little & Rubin, 2002). Depending on the mechanism generating the missing data, the literature describes different techniques to deal with this eventuality (Allison, 2009; Enders, 2010). This choice of technique, in dealing with missing data, can influence the conclusions reached regarding change over time (Little & Rubin, 2002; Wothke, 2000). Next to this, methodologists, in social sciences, recently described techniques stemming from biostatistics for MNAR missing data (Enders, 2011; Hedeker & Gibbons, 1997). In light of this, it was advocated that researchers conduct a sensitivity analysis to gauge the stability of the models' results to missing data techniques, assuming different missing data mechanisms (Enders, 2011; Molenberghs & Fitzmaurice, 2009).

Nevertheless, there is a paucity of practical examples of such sensitivity analysis in educational sciences. Moreover, there are two families of MNAR techniques:

selection models on the one hand and PM models on the other hand. Former studies in the educational domain have focussed (predominantly) on selection models (Foster et al., 2004; Xu & Blozis, 2011). There are few practical examples of sensitivity analysis using PM models. Taking the case of change in students' learning strategies during higher education, this study aims to provide a tutorial example of conducting a sensitivity analysis using various missing data techniques, among which PM models.

If the missing data technique influences the results, and more importantly, the substantive conclusions drawn from them, this is clearly of interest to educational researchers estimating change over time. In the case of missing data, which technique (or combination of techniques) can be recommended as the 'best' choice? For this reason, insight into the missing data mechanisms, and into sensitivity analysis, is also important for educational researchers.

6.1.1 Mechanisms generating missingness

The literature discerns three mechanisms by which missing data could arise (Little & Rubin, 2002). Though they are applicable to all data collections with missing data, we illustrate them here for longitudinal data. For a more general and technical description of the mechanisms, we refer to Graham (2009) and Enders (2010). Given that the mechanisms abbreviations are easily mistaken one for another, Table 6.1 provides a short summary.

Firstly, data may be MCAR. This is the case when "the missing time points are a random sample of all time points and the dropouts are a random sample of all participants" (Raudenbush, 2001a, p. 517). An example from the learning strategies domain could be that a student misses a data collection due to illness (e.g. the flu). Yet, as several authors pointed out, assuming that longitudinal data were *solely* MCAR was stringent and unlikely to hold up in practice (Enders & Bandalos, 2001; Raudenbush, 2001a). For example, concerning longitudinal research on student learning strategies, there is an extensive body of research on the influence of student learning strategies on academic achievement (e.g., Richardson et al., 2012; Vermunt, 2005), which suggests that students, in a non-delayed study trajectory, are unlikely to be a random sample of those starting higher education. This suggests that data will probably not be MCAR.

A second mechanism, generating missing data, holds when the probability of missingness is related to one or several variables in the study, such as the score at a previous wave. This is labelled with the - very confusing - term MAR (Collins, Schafer, & Kam, 2001). An example would be when students, scoring higher on surface processing of the first wave, were found to have a higher chance of dropping out of higher education, and therefore are found to be absent at the second wave.

Table 6.1: Detail on abbreviations regarding missingness

Abbreviation	Full term	Probability of missingness is related to	Technique
MCAR	Missing Completely at Random	chance	1. Listwise deletion (LD)
MAR	Missing at Random	variable(s) in the study (e.g., score at previous wave, demographic characteristic)	2. Maximum Likelihood (ML) 3. Multiple Imputation (MI) 4. ML with auxiliary variables (MLaux) 5. MI with auxiliary variables (MIaux)
MNAR	Missing Not at Random	the unobserved change over time (e.g., change in scores between waves)	Pattern mixture (PM) models: 6. Hedeker & Gibbons (H&G); 7.-9. Models with identifying restrictions

Thirdly, the chance of being missing can depend on the missing value (outcome-related missing data or MNAR). This would be the case if students, who decreased their deep processing from wave one to wave two, were more prone to drop out of higher education prior to wave two. The – unobserved - change over time in deep processing then predicts the chance of being missing. The difference between the MAR and MNAR mechanisms lies in the assumption of whether or not variables related to missingness were present in the collected data. With MAR, it is assumed that missingness is solely related to variables in the dataset (e.g., value at time 1). When this relationship is accounted for, there is no further relationship between the missing values and the chance of being missing. With MNAR, on the other hand, the change over time, leading to the missing data, is assumed to be unobserved.

In non-simulated longitudinal data all three mechanisms generating missing data may be present (Yoo, 2009). Some missing data can be due to reasons in line with MCAR, whilst other missing data are caused by MAR or MNAR mechanisms. Whether or not missing data are related solely to MCAR can be falsified using independent samples *t*–test. It contrasts the students' scores of the longitudinal group, at a certain wave, to their peers with missing data. When there is a significant mean difference, the MCAR assumption is rejected. Consequently, missing data is also MAR. Note that the absence of a significant mean difference does not prove the MCAR assumption. Missing data may still be due to MAR, but perhaps due to low statistical power (e.g., small N) independent samples *t*-test did not detect significant differences.

In contrast to the MCAR assumption, neither the MAR nor the MNAR assumption can be falsified, due to the fact that the answer lies partly within the absent data (Peugh & Enders, 2004; Xu & Blozis, 2011). For this reason, sensitivity analysis is suggested.

6.1.2 Sensitivity analysis

When outcome-related missing data (MNAR) is plausible, it is recommended to gauge the sensitivity of the parameter estimates to the mechanism for missing data (Enders, 2011; Xu & Blozis, 2011). The results from techniques, assuming MAR and MNAR, should be compared Prior to discussing missing data techniques assuming MAR or MNAR, we will first discuss listwise deletion (LD) - which assumes MCAR - given that this missing data technique is currently predominant in educational research (Peugh & Enders, 2004). We aimed to provide a non-technical review focussed on longitudinal data. For a full review and technical detail, please see the references cited in the following sections.

A. Technique assuming MCAR: listwise deletion

Most longitudinal studies, on the change in students' learning strategies during higher education, rely upon LD. Only those cases, with values at each of the three (or more) waves, are retained in the dataset (Busato et al., 1998; Van der Veken et al., 2009; Zeegers, 2001). Sometimes, this results in a very small longitudinal group (e.g., N=26, Busato et al., 1998; N=43, Zeegers, 2001). Two studies tested the tenability of the MCAR assumption. In the study of Busato et al. (1998) the MCAR assumption was not rejected, whilst Van der Veken et al.

(2009) found differences between responding students and their non-responding peers.

Even if the MCAR assumption is not disproven, LD is judged to be suboptimal due to the lower power caused by the reduction in sample size (Allison, 2009; Peugh & Enders, 2004). The MCAR technique does not use the available data efficiently (Wothke, 2000). Therefore, methodologists and the APA Task Force on Statistical Inference strongly advised against LD (Enders, 2010; Little & Rubin, 2002), judging it "among the worst methods for practical applications" (Wilkinson & Task Force on Statistical Inference, 1999, p. 598).

B. *Techniques assuming MAR: maximum likelihood, multiple imputation and including auxiliary variables*

Maximum likelihood (ML) estimates the parameters, which are most likely to have produced the sample data, by trying out different values for these parameters. To obtain ML parameter estimates, the EM algorithm is a commonly used approach. EM takes the form of two steps (Expectation or E-step and Maximization or M-step), which inform one another and are repeated after each other (i.e., iterative procedure, Peugh & Enders, 2004). With data containing missing values, the E-step's goal is to estimate the missing values using information from the observed values. To do this, information on the mean vector and the complete variables' covariance matrix is used to generate regression equations that predict incomplete values. Next, in the M-step, an updated mean vector and covariance matrix is calculated; this relies upon both the available data and the estimated data for the missing values. In the next E-step, with this updated mean vector and covariance matrix, the missing values are estimated anew and lead to another update of the mean vector and covariance matrix (M-step). This process stops when the mean vector and covariance matrix hardly alters between two successive M-steps (i.e., convergence is reached).

As shown in Table 6.1, a second technique, assuming MAR, is multiple imputation (MI), which involves three phases. In the imputation phase, m (e.g., 100) complete datasets are generated by filling in the missing values with different plausible estimates. In the analysis phase, analysis is done on each of the m complete datasets. In the third phase, the parameter estimates and the standard errors are pooled. Parameter estimates are calculated as the mean over the m datasets, whilst their standard errors are computed by taking into account both the variance within datasets (within variance) and between datasets (between variance, Allison, 2009).

In both the ML and MI techniques, auxiliary variables can be included (MLaux and MIaux). They can be conceptualised as variables that are included in order to estimate (ML) or fill in (MI) missing values in a more informed or accurate way. Thereby, by predicting some of the missingness, MNAR missingness may be rendered into MAR (Allison, 2009; Yoo, 2009).

Two types of variables are of interest as auxiliary variables (Collins et al., 2001; Yoo, 2009). Firstly, variables that predict missingness can be informative. If, for example, girls are more likely to participate, gender can be used. Secondly, variables correlating moderately to strongly with the variables under study are of interest. If deep approach and grade point averages are correlated, the latter can be included when estimating or imputing the missing data for the deep approach variable.

C. Techniques assuming MNAR: pattern mixture (PM)

There are two families of MNAR models (Enders, 2011; Molenberghs & Fitzmaurice, 2009). The PM models first divide the sample into subgroups depending on missing data patterns (e.g., a group with complete information, a group with only information at the first wave). Next, for each of the subgroups, the parameter estimates are calculated. Finally, the results of the different models are put together. A second family consists of the selection models which estimate the probability of missingness and the parameters simultaneously in one model.

Both families of models rely on a number of untestable assumptions. For the selection models, small departures from the multivariate normality assumption can have a serious bias on the results (Demirtas & Schafer, 2003; Foster et al., 2004). For the PM models, restrictions have to be imposed in order to allow the models to be estimated for all subgroups (Molenberghs & Fitzmaurice, 2009). Yet, these last assumptions are explicit and different restrictions can be used to assess the sensitivity of the results to these models (Enders, 2011; Foster et al., 2004). For these reasons, we focused our study on PM models. More specifically, we selected four PM models.

The first PM model was the Hedeker and Gibbons (H&G) model, which assesses the parameter estimates for respondents with complete data (subgroup 1) and for respondents with incomplete data (subgroup 2). Subsequently, by taking the proportion of each subgroup into account, using the weighted average, the population parameters are calculated. An advantage of the H&G model is that longitudinal data can usually be divided into the two subgroups without the number per subgroup becoming too small. A drawback is that all respondents with incomplete data are treated alike. This may not always make sense intuitively; for example, students dropping out after the first year may be a different type of student than those dropping out after the second year.

The other three PM models rely on more subgroups. For example, in a three wave study, three subgroups can be discerned: students with complete data (1); those who go missing after the second wave (2); and students who go missing after the first wave (3). Yet, students in subgroup 3 have only one data point. In order to estimate their change over time, three types of identifying restrictions can be put in place (hence, three PM models, Demirtas & Schafer, 2003). In the *complete case* restriction, the parameter estimate of the group providing complete data (subgroup 1) is used: the change over time for subgroup 3 is restricted to (or put equal to) the change over time for the complete cases (subgroup 1). The *neighbouring case* restriction equals the parameter estimate for subgroup 3 to that of subgroup 2. The *available case* option consists of using the weighted average of the respective parameter estimates of subgroups 1 and 2. Once parameter estimates have been calculated for each of the three subgroups, the population parameter estimate is derived by using the weighted average.

6.1.3 This study

This study aims to provide a tutorial example of sensitivity analysis. For this, a non-simulated dataset on the change in students' learning strategies during higher education is used. Thus, comparable to educational researchers confronted with missing data, the mechanism causing the missing data (MCAR, MAR or MNAR) is unknown. This non-simulated data allows us to provide a genuine example of using sensitivity analysis in educational research, and illustrate the guidelines in reporting the results from this analysis.

The results from techniques assuming MAR (ML, MI, MLaux and MIaux) were checked against those assuming MNAR (PM models: H&G model, complete, neighbouring and available case restriction model). Next to this, we included LD, assuming MCAR. Although strongly cautioned against (Little & Rubin, 2002; Marsh & Hau, 2007; Wilkinson & Task Force on Statistical Inference, 1999), LD is used predominantly in studies on the change in students' approaches to learning over time and, more generally, in education research (Peugh & Enders, 2004). Therefore, we chose to include this technique next to the models assuming MAR and MNAR.

The research questions were:

1. In assessing change over time in student learning strategies, do results differ according to the missing data technique (assuming MCAR, MAR and MNAR) adopted?
2. Do the different results lead to substantively different conclusions regarding the growth over time?

These research questions are answered in two studies. When confronted with missing data, researchers relying on LD sometimes - out of caution - only generalize to participants who persist (e.g., students progressing normally throughout higher education). However, the analyses are done on those students with complete data; students with non-response are ignored. When doing so, it is implicitly assumed that taking missingness, due to non-response into account or not, does not influence the results and the substantive conclusions drawn from them. In study 1 we tested this assumption: can results and conclusions from students with complete data (using LD) be generalised to all students progressing normally during higher education?

A second option often taken by educational researchers is to ignore missingness (using LD), but nevertheless generalizing to all participants (e.g., students in higher education). In doing so, it is assumed that whether missingness, due to attrition or non-response is taken into account, does not influence the conclusions reached. In study 2, we examined the effect of both non-response and attrition by using all available data points. Here, we expect larger differences in the results and conclusions between the models compared to in study 1, given that studies indicate that attrition cannot be assumed to be MCAR (e.g., Richardson et al., 2012), but MAR and MNAR are more likely.

6.2 Method

6.2.1 Design and instrument

The research took place in a Belgian university college in which one cohort of students was followed up. In March of the first academic year (from September to June), first-year students participated in the research. The same cohort was again questioned during May in both the second and third year of study.

Table 6.2: Three learning strategy scales of the ILS-SV; number of items; item example (translated from Dutch); and reliability estimates

Scales	Items	Item example	α°
Memorizing	4	I learn definitions by heart and as literally as possible.	.68-.71
Lack of regulation	4	I confirm that I find it difficult to establish whether or not I have sufficiently mastered the course material.	.68-.73
Analysing	4	I study each course book chapter point by point and look into each piece separately.	.66-.70

° the lowest and highest α obtained for each of the three waves is given

Students' learning strategies consist of cognitive processing and regulation activities and are mapped using the Inventory of Learning Styles – Short Version (ILS-SV, Donche & Van Petegem, 2008). Three of the seven learning strategy scales were selected from a tutorial perspective: the memorizing, lack of regulation and analysing scales presented an array of possible outcomes from sensitivity analysis, which researchers may encounter in practice. By selecting a small number of scales, the results and suggestions for practice can be presented in detail. Table 6.2 provides for the three scales: the number of items, an example item, and the reliability estimates.

6.2.2 Latent growth analysis

In order to assess growth over time, we relied upon latent growth analysis. This allows the estimation of MNAR models more easily than multilevel analysis (Enders, 2010). Figure 6.1 depicts such a model that estimates the average growth in the manifest scale scores by an intercept and a slope. The intercept signifies the initial value for the scale, whilst the slope indicates whether or not, on average, there is an increase or decrease in the scale scores per unit of time (here, 12 months).

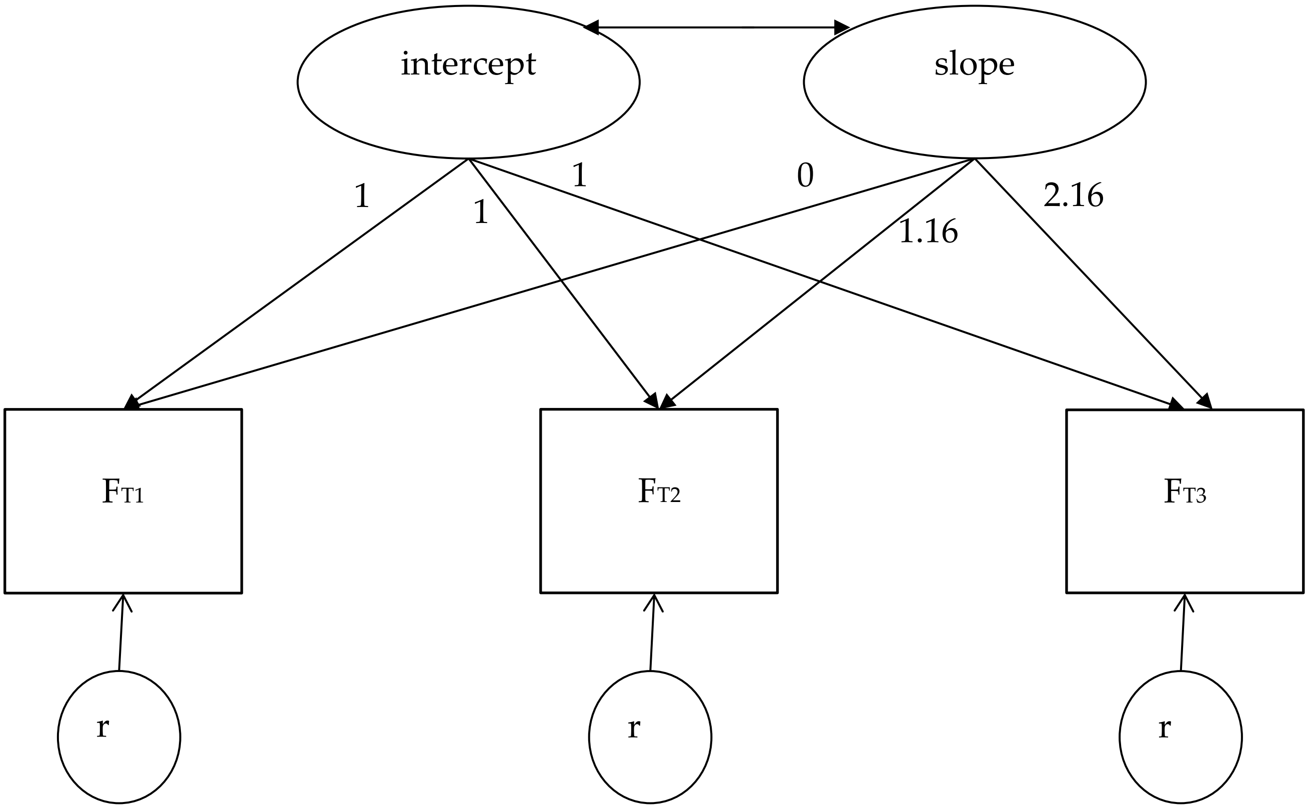


Figure 6.1: Latent growth model

Three parameters are estimated next to the average growth trajectory. Firstly, the intercept variance parameter expresses whether or not students varied significantly in their initial level as measured by a learning strategy scale. Secondly, the variance in slopes indicates whether the students followed the general trend or deviated from one another. Thirdly, if both the intercept and slope variance prove significant, the covariance becomes a point of interest. This covariance indicates whether or not, on a learning strategy scale, the students' initial scores were related to their change over time.

6.2.3 Missing data and options to handle them

Table 6.3 provides detail on the missing data in the sample. There is a considerable amount of attrition. Of the cohort under study, 1,355 students were registered in the first year and 410 (30.3%) of those students continued in a non-delayed study trajectory onto the third year. Due to unrestricted entrance into higher education in Belgium, it is commonplace that a large number of students stop during the first year or fail their exams (e.g., Germeijs & Verschueren, 2007). Next to attrition, there was wave non-response. Given that students were questioned during lecture slots, the response rates were adequate: 76.1% of the eligible students participated in wave 1, descending to 66.6% in wave 3 (see Table 6.3). Thirdly, there was a small number of item non-response. Scale scores were computed only if the student answered each of the four items for a learning strategy scale. Therefore, item non-response was treated as wave non-response. For example, at the first wave 6 students did not complete all items for the analysing scale (see Table 6.3).

Table 6.3: Registration, participation and response rate per measurement wave

	Wave 1	Wave 2	Wave 3
Number of registered students	1355	616	410
Number of respondents	1031	442	279
Response rate (%)	76.1	70.5	66.6
Number of respondents without item non-response$_{\text{Memorizing and Lack of regulation}}$	1029	442	278
Number of respondents without item non-response$_{\text{Analysing}}$	1025	440	275

In total 21.8% (=225/1029) of the students, who participated in the first wave, provided complete information at each of the three waves (21.6% for the analysing scale, 222/1025)[12]. This percentage was in line with (e.g., 21.5%; Zeegers, 2001), or better than, other studies on the change in student learning strategies during higher education (7.5%, Busato et al., 1998; 6.5%, Jackling, 2005).

A. Study 1

For study 1, examining the effects of non-response, we selected the students from whom administrative data informed they had followed a non-delayed trajectory during higher education (i.e., progressed normally throughout the three years of higher education, here, N=410). Of these students, 225 provided complete data at each of the three waves for the memorizing and lack of regulation scale, whilst 222 provided complete data for the analysing scale (further named as the longitudinal group). Fifteen students did not participate in any of the three waves and, therefore, were omitted from the analysis. Three hundred and ninety-five students provided data at one of the three waves.

[12] One may note that this is less than the longitudinal group of 245 retained in chapters 3 and 4. For this sixth chapter, administrative data were relied upon to discern drop-out patterns (see further). In doing so, it was noted that some students had responded to the questionnaire, without being enrolled in that year. This could be due to students for example having not succeeded in all the courses of the second year, following a more individual study trajectory in which they would re-take a number of courses from the second year and take some courses in the third year as well. These students were included in chapters 3 and 4 given they responded at all waves, but are excluded from analysis in this chapter given they do not meet the administrative requirements for a non-delayed study trajectory.

B. *Study 2*

For study 2, examining the effect of both non-response and attrition, we made use of all available data. The number of students providing data for at least one wave for the analysing scale was 1,071, whilst 1,072 did so for the memorizing and lack of regulation scale. The 283 students, who did not provide data at any of the three waves, were excluded from the analysis.

6.2.4 *Plan of analysis*

The plan of analysis was alike for both studies 1 and 2. First, MCAR assumption was falsified using independent samples *t*-tests in SPSS. Second, latent growth analysis is done in Mplus 6.1 using nine missing data techniques (see Table 6.1). For the memorizing scale in study 2, an annotated syntax for each of these latent growth analyses is included in section 6.5. Sample data to reproduce the findings are available upon request.

Assuming MCAR, we estimated the latent growth model using LD ($N_{study1\&2}$=225 for memorizing and lack of regulation; $N_{study1\&2}$=222 for analysing). We ran the other eight models on the sample of students providing data on at least one wave (N_{study1}=395; N_{study2}=1071 for analysing and 1072 for memorizing and lack of regulation). We estimated four techniques assuming MAR: ML (through the EM algorithm, Muthén, Asparouhov, Hunter, & Leuchter, 2011); MI; MLaux; and MIaux. For the MI and MIaux models, we opted for 100 imputed datasets, given the large percentage of missing data in study 2. Moreover, a higher number of imputed dataset could increase the stability of the estimates and, since the latent growth model required only a short computational time, there was no drawback in including more datasets (Enders, 2010; Graham, Olchowski, & Gilreath, 2007).

Table 6.4: Explained variance in the chance of missingness by the auxiliary variables (Nagelkerke R^2; in %)

	Study 1			Study 2		
	Wave 1	Wave 2	Wave 3	Wave 1	Wave 2	Wave 3
$\text{Missing}_{\text{Memorizing}}$	26.7	7.7	10	20.0	7.3	18.4
$\text{Missing}_{\text{Lack of regulation}}$	20.0	7.7	10	20.0	7.3	18.4
$\text{Missing}_{\text{Analysing}}$	19.4	7.8	10.1	20.1	7.4	18.4

The following administrative data were available as auxiliary variables: gender; prior education (general, technical or vocational); study track in higher education; whether or not students had started the first year in that university college anew; whether or not they had followed a non-delayed study trajectory; and the grade point average for each year. Good auxiliary variables predict the chance of being missing or are correlated with the variables under study (Collins et al., 2001). To examine the former, logistic regression was used to determine the variance explained by the auxiliary variables, in whether students were missing at a wave or not. The results are given in Table 6.4. To examine the latter, regression was used to provide the explained variance in memorising, lack of regulation and analysing by the auxiliary variables (see Table 6.5).

Table 6.5: Explained variance of the variables under study by the auxiliary variables (R^2; in %)

	Study 1			Study 2		
	Wave 1	Wave 2	Wave 3	Wave 1	Wave 2	Wave 3
Memorizing	9.3	4.8	7.1	10.7	8.0	8.3
Lack of regulation	13.5	15.3	20.9	14.4	17.3	20.4
Analysing	10.3	10.9	14.0	10.5	9.6	12.5

Table 6.6: Missing data patterns for the memorizing scale for study 1

Pattern	Wave1	Wave2	Wave3	Number of students	% of students
1	O	O	O	225	54.9
2	O	O	M	62	15.1
3	O	M	M	42	10.2
4	M	O	O	6	1.5
5	O	M	O	35	8.5
6	M	O	M	13	3.2
7	M	M	O	12	2.9
8	M	M	M	15	3.7

Note: *O* = data observed; *M* = data missing

For both study 1 and 2, the auxiliary variables do indeed explain a part of the variance. Following Cohen's (1988) rules for interpretation of the R^2 (see Table 6.5)[13], the explained variance ranges from a small to a large effect (4.8%-20.9%). However, though research has shown that good auxiliary variables are characterised by a larger explained variance (49%-81%, Collins et al., 2001; Johnson & Young, 2011), no harm was found in including auxiliary variables little related to the variable under study (Collins et al., 2001).

[13] We note that the Nagelkerke R^2 from Table 6.4 cannot be interpreted in absolute terms (Cohen, Cohen, West, & Aiken, 2003).

The techniques, assuming MNAR, consisted of the H&G model and three models with identifying restrictions. For the models with identifying restrictions, the main hurdle for a researcher is to decide how to form subgroups. Looking at the missing data patterns for the memorizing scale as shown in Table 6.6, usually many possible patterns are suggested, which may be too small in number for reliable analysis (e.g., pattern 4). The key is then to combine subgroups. Note that the group with three missing data points (pattern 8) is left out of the analysis. For study 1, investigating the effect of non-response given that only students persisting in their studies were included, there are a number of options.

1. Subgroups on the last wave for which students were observed (wave 3: patterns 1, 4, 5 and 7; wave 2: patterns 2 and 6; wave 1: pattern 3).
2. Subgroups on intermittent missingness (patterns 1 to 3 as 3 different subgroups, grouping patterns 4 to 7 as the intermittent subgroup), which is of interest when one is keen to know whether students with intermittent missingness differ from their peers.
3. Subgroups on the number of missing data points (none: pattern 1; 1 missing data point: patterns 2, 4 and 5; and two missing data points: patterns 3, 6 and 7). Using this division into subgroups sheds light on whether students with more missing data due to non-response differ from their peers.

Keeping the research questions in mind when choosing between the options is clearly important, as is the size of the subgroups. For study 1, the H&G model contrasted students with complete data, with those having incomplete data. For the models with identifying restrictions we chose the third option, given that the effect of missingness due to non-response was under study. The three groups were thus: students with complete data (subgroup 1); students with one non-response (subgroup 2); and those with two missing data points (subgroup 3). For the third subgroup, identifying restrictions had to be put into place in order for the growth model to be estimated. In the complete case restriction, the change over time for the third group was set equal to subgroup 1 (those with complete data, see 6.5 for the annotated syntax). In the neighbouring case restriction, their change over time was restricted to that of subgroup 2. For the available case restriction, the weighted average of the growth estimates for subgroups 1 and 2 is used.

In study 2, we relied upon administrative data to discern subgroups of attrition. For the H&G model, students progressing normally throughout their three years of study (N=395) were contrasted with their peers who did not progress normally (N=677). For the identifying restriction models, we discerned three subgroups: students in a non-delayed trajectory (1, N=395); those registered up to second year (2, N=184); and those registered only in the first year (3, N=493). Here, it is evident that the growth of subgroup 3 cannot be estimated. Therefore, in the complete case, the growth for subgroup 3 was constrained to the growth of subgroup 1. In the neighbouring case, it was restrained to the growth of subgroup 2. Lastly, with the available case restriction, we used the weighted average of subgroups 1 and 2.

6.3 Results

6.3.1 Study 1

The first study examined the effects of excluding or including students with non-response. Prior to this, using independent samples *t*-tests, the MCAR assumption was falsified. For the memorizing and analysing scale, no significant differences were detected between students from the longitudinal group and those with incomplete data. For the lack of regulation scale, students, from the longitudinal group, scored significantly lower in the second wave than their colleagues ($t(116.12)=3.567$, $p<.01$, Cohen's $d=.36$). For this scale, the MCAR assumption was thus rejected.

Table 6.7 presents the results of the latent growth models using nine different missing data techniques for the memorizing, lack of regulation, and analysing scales. Looking at the results for the memorizing scale, the analysis, using LD, suggests that, during higher education, students reduce their reliance on this processing strategy. We noted variance in the intercept but not in the slopes. Next, models assuming MAR (ML, MI, MLaux and MIaux) indicated very similar estimates on the average and differential growth. The results from the models assuming MNAR were in line as well. For the memorizing scale, the results from the nine missing data techniques were thus very comparable.

Table 6.7: Parameter estimates and standard errors for study 1's growth models

	Intercept	Slope	Intercept variance	Slope variance	Covariance
Memorizing°					
MCAR					
Listwise deletion (LD)	3.318 (.054)***	-.083 (.027)**	.409 (.087)***	.032 (.033)	
MAR					
Maximum Likelihood (ML)	3.319 (.043)***	-.079 (.023)**	.428 (.081)***	.031 (.032)	
Multiple Imputation (MI)	3.318 (.044)***	-.078 (.023)**	.432 (.083)***	.032 (.033)	
ML with auxiliary variables (MLaux)	3.317 (.043)***	-.073 (.023)**	.438 (.082)***	.032 (.033)	
MI with auxiliary variables (MIaux)	3.318 (.044)***	-.078 (.023)**	.432 (.083)***	.032 (.033)	
MNAR					
Hedeker & Gibbons (H&G)	3.317 (.043)***	-.075 (.025)**	.419 (.057)***	.022 (.018)	
Complete Case	3.319 (.043)***	-.078 (.023)**	.419 (.055)***	.022 (.019)	
Neighbouring Case	3.317 (.043)***	-.075 (.025)*	.419 (.055)***	.022 (.019)	
Available Case	3.318 (.043)***	-.077 (.023)**	.419 (.055)***	.022 (.019)	
Lack of regulation°					
MCAR					
Listwise deletion (LD)	2.569 (.053)***	-.135 (.025)***	.350 (.077)***	.016 (.030)	
MAR					
Maximum Likelihood (ML)	2.638 (.043)***	-.138 (.022)***	.457 (.077)***	.048 (.029)	
Multiple Imputation (MI)	2.641 (.044)***	-.141 (.023)***	.459 (.077)***	.049 (.029)	
ML with auxiliary variables (MLaux)	2.638 (.044)***	-.138 (.023)***	.484 (.078)***	.062(.029)*	-.045 (.039)
MI with auxiliary variables (MIaux)	2.642 (.044)***	-.142 (.023)***	.488 (.079)***	.062 (.030)*	-.047 (.041)
MNAR					
Hedeker & Gibbons (H&G)	2.640 (.044)***	-.131 (.025)***	.466 (.057)***	.036 (.017)*	-.031 (.024)
Complete Case	2.640 (.044)***	-.136 (.024)***	.463 (.052)***	.035 (.016)*	-.030 (.021)
Neighbouring Case	2.638 (.044)***	-.132 (.027)**	.463 (.052)***	.036 (.016)*	-.033 (.021)
Available Case	2.640 (.044)***	-.136 (.024)***	.463 (.052)***	.035 (.016)*	-.030 (.021)

Table 6.7 (continued): Parameter estimates and standard errors for study 1's growth model

	Intercept	Slope	Intercept variance	Slope variance	Covariance
Analysing°°					
MCAR					
Listwise deletion (LD)	3.060 (.053)***	.009 (.026)	.340 (.079)***	.033 (.031)	
MAR					
Maximum Likelihood (ML)	3.056 (.041)***	.013 (.022)	.357 (.069)***	.035 (.028)	
Multiple Imputation (MI)	3.059 (.040)***	.010 (.022)	.358 (.070)***	.037 (.029)	
ML with auxiliary variables (MLaux)	3.062 (.041)***	.010 (.022)	.359 (.070)***	.036 (.028)	
MI with auxiliary variables (MIaux)	3.065 (.041)***	.006 (.022)	.354 (.070)***	.037 (.029)	
MNAR					
Hedeker & Gibbons (H&G)	3.054 (.041)***	.017 (.024)	.379 (.049)***	.031 (.016)	
Complete Case	3.058 (.041)***	.014 (.022)	.379 (.050)***	.031 (.017)	
Neighbouring Case	3.056 (.041)***	.018 (.024)	.379 (.050)***	.031 (.017)	
Available Case	3.058 (.041)***	.014 (.023)	.379 (.050)***	.031 (.017)	

*** $p<.001$; ** $p<.01$; * $p<.05$; ° For the LD model: N= 225, for all other models: N=395; °° For the LD model: N= 222, for all other models: N=395

Note: The auxiliary variables were gender; prior education (general, technical or vocational); study track in higher education; whether or not students had started the first year in that university college anew; whether or not they had followed a non-delayed study trajectory; and the grade point average for each year.

For the lack of regulation scale, the LD estimates suggest a significant decrease over time. Regarding the variances, we noted significant intercept variance, whilst the results indicated no differential growth over time. The MAR models estimated the intercept somewhat larger than the LD model, but the 95% confidence intervals (CI) overlap[14]. The slope parameter was comparable to the one from the listwise deleted sample. Regarding differential growth, the MAR models estimated a larger variance in the intercept, although they reached the significance level only for the MLaux and MIaux (CI LD 0.29-0.42; CI MLaux 0.43-0.53; CI MIaux 0.43-0.56)[15]. The models including auxiliary variables, also suggested a significant slope variance. Examining the MNAR models revealed that the intercept and slope were comparable to the MAR models. In line with MIaux, the MNAR models indicated slope variance. In summary, all models confirmed the general decreasing trend in lack of regulation. However, the models differed in the magnitude of the intercept variance and in their assessment of differential growth over time.

The results for the analysing scale were straightforward. The MCAR, MAR and MNAR models provided comparable estimates for the intercept, slope and the two variance parameters. For students progressing normally through higher

---

14 The 95% CI for an intercept is not provided in the Mplus output but is easily calculated as follows: intercept ± 1.96*standard error of the intercept. The standard error of the intercept is calculated as the √intercept variance/√number of respondents in the dataset. For example, for the LD sample the standard error of the intercept= √0.350/√225=0.0394. The CI is then 2.569 ± 1.96*0.0394.

15 The 95% CI around variance parameters is not provided in the Mplus output but can be computed using for example R. It requires information on the cumulative probability of the Chi² distribution, which can be in R with the qchisq function. Calculate the variance's lower limit as ((N-1)*variance)/qchisq(.975,N-1); the upper limit at ((N-1)*variance)/qchisq(.025,N-1). For example, for the MIaux results the lower limit is (394*0.488)/450.88=0.427 and the upper limit is (394*0.488)/340.89=0.564.

education, all models confirmed a constant trend over time. Students were found to vary significantly in their degree of analysing at the first wave, but not in their change over time.

6.3.2 Study 2

In study 2, to investigate the effect of both non-response and attrition on the results, we made use of all available data. Students who dropped out of their studies are thus included in the analysis. First, the MCAR assumption was verified. The results of independent samples *t*-tests for the memorizing scale indicated no significant differences between students from the longitudinal group and their peers with incomplete data. Concerning the lack of regulation scale, students, from the longitudinal group, scored significantly lower in both the first and second waves ($t(1027)=5.973$, $p<.001$, Cohen's $d=.45$; $t(440)=4.950$, $p<.001$, Cohen's $d=.47$). Students from the longitudinal group also scored significantly higher on analysing in both the first and second waves ($t(1023)=-3.295$, $p<.01$, Cohen's $d=.25$; $t(438)=-2.680$, $p<.01$, Cohen's $d=.26$). Therefore, we rejected the MCAR assumption for the lack of regulation and analysing scale.

Table 6.8: Parameter estimates and standard errors for study 2's growth models

	Intercept	Slope	Intercept variance	Slope variance	Covariance
Memorizing°					
MCAR					
Listwise deletion (LD)	3.318 (.054)***	-.083 (.027)**	.409 (.087)***	.032 (.033)	
MAR					
Maximum Likelihood (ML)	3.278 (.027)***	-.063 (.019)**	.424 (.069)***	.015 (.030)	
Multiple Imputation (MI)	3.282 (.027)***	-.067 (.019)***	.419 (.072)***	.014 (.029)	
ML with auxiliary variables (MLaux)	3.276 (.027)***	-.076 (.029)**	.421 (.069)***	.014 (.030)	
MI with auxiliary variables (MIaux)	3.282 (.030)***	-.070 (.024)**	.413 (.069)***	.014 (.030)	
MNAR					
Hedeker & Gibbons (H&G)	3.274 (.027)***	-.002 (.036)	.453 (.040)***	.029 (.018)	
Complete Case	3.274 (.027)***	-.058 (.021)**	.453 (.040)***	.029 (.019)	
Neighbouring Case	3.274 (.027)***	-.003 (.038)	.453 (.040)***	.029 (.019)	
Available Case	3.274 (.027)***	-.041 (.024)	.453 (.040)***	.029 (.019)	
Lack of regulation°					
MCAR					
Listwise deletion (LD)	2.569 (.053)***	-.135 (.025)***	.350 (.077)***	.016 (.030)	
MAR					
Maximum Likelihood (ML)	2.869 (.026)***	-.182 (.020)***	.440 (.062)***	.051 (.027)	-.025 (.036)
Multiple Imputation (MI)	2.874 (.027)***	-.187 (.024)***	.453 (.060)***	.054 (.026)*	-.033 (.035)
ML with auxiliary variables (MLaux)	2.868 (.026)***	-.104 (.031)**	.464 (.061)***	.063 (.027)*	-.025 (.036)
MI with auxiliary variables (MIaux)	2.891 (.028)***	-.149 (.029)***	.522 (.067)***	.082 (.027)**	-.061 (.038)
MNAR					
Hedeker & Gibbons (H&G)	2.890 (.026)***	-.143 (.037)***	.403 (.036)***	.030 (.017)	
Complete Case	2.865 (.026)***	-.130 (.021)***	.400 (.035)***	.030 (.016)	
Neighbouring Case	2.865 (.026)***	-.113 (.038)**	.400 (.035)***	.030 (.016)	
Available Case	2.865 (.026)***	-.125 (.024)***	.400 (.035)***	.030 (.016)	

(table continues)

Table 6.8 (continued): Parameter estimates and standard errors for study 2's growth models

	Intercept	Slope	Intercept variance	Slope variance	Covariance
Analysing°°					
MCAR					
Listwise deletion (LD)	3.060 (.053)***	.009 (.026)	.340 (.079)***	.033 (.031)	
MAR					
Maximum Likelihood (ML)	2.890 (.024)***	.049 (.020)*	.350 (.058)***	.033 (.026)	
Multiple Imputation (MI)	2.901 (.027)***	.039 (.021)	.338 (.058)***	.029 (.027)	
ML with auxiliary variables (MLaux)	2.890 (.025)***	-.018 (.030)	.358 (.058)***	.036 (.027)	
MI with auxiliary variables (MIaux)	2.904 (.028)***	-.034 (.035)	.343 (.061)***	.031 (.029)	
MNAR					
Hedeker & Gibbons (H&G)	2.890 (.025)***	.039 (.034)	.360 (.033)***	.033 (.016)*	-.006 (.019)
Complete Case	2.890 (.025)***	.019 (.021)	.359 (.035)***	.033 (.017)*	-.006 (.019)
Neighbouring Case	2.890 (.025)***	.028 (.036)	.359 (.035)***	.033 (.017)*	-.006 (.019)
Available Case	2.890 (.025)***	.021 (.023)	.359 (.035)***	.033 (.017)*	-.006 (.019)

*** $p<.001$; ** $p<.01$; * $p<.05$; ° For the LD model: N= 225, for all other models: N=1071; °° For the LD model: N= 222, for all other models: N=1071

Note: The auxiliary variables were gender; prior education (general, technical or vocational); study track in higher education, whether or not students had started the first year in that university college anew; whether or not they had followed a non-delayed study trajectory; and the grade point average for each year.

Table 6.8 presents the results from the sensitivity analyses. For the memorizing scale, the models, assuming either MCAR or MAR, indicated a significant decrease over time, combined with variance in intercepts but not in slopes. However, the models, assuming MNAR, did not confirm a declining trend over time. The H&G model estimated that the general trend was insignificant due to the fact that students, who dropped out of their studies at a certain time or fell behind, were found to remain constant on this learning strategy scale (*b*=.043, *se*=.055, *p*=.43; not in table). The estimates from the neighbouring and available case models also suggested absence of change over time. In summary, the models, relying on different assumptions regarding missingness, disagreed on whether or not there was average growth.

Concerning the lack of regulation scale, the intercept was estimated to be significantly higher for the MAR and MNAR models compared to the MCAR model (CI LD 2.44-2.65; CI ML 2.83-2.91). The estimate of the slope did not differ significantly between the models. Compared to the MCAR model, and although significant only for the MLaux and MIaux models (CI LD .29-.42; CI MLaux .43-.49; CI MIaux .49-.55), the intercept variance was estimated to be higher in the MAR and MNAR models. Whilst the MCAR model did not detect slope variance, the estimates for the ML model were at the verge of significance (*var slope*=.051, *se*=.027, *p*=.054) and those of the MI, MLaux and MIaux models reached significance (see Table 6.8). However, the MNAR models did not confirm this differential growth. Consequently, the MCAR models differed from the MAR models on the intercept and on the slope variance, whilst the MNAR and MAR models disagreed regarding this last parameter.

For the third scale, analysing, we noted that the MAR and MNAR models estimated the intercept significantly lower compared to the MCAR model (CI LD: 2.98-3.13; CI ML: 2.85-2.93). The trend over time was estimated to be constant using the LD, MI, MLaux and MIaux techniques and MNAR models. On the other hand, the ML estimates suggested an increase over time. Concerning the intercept variance, we found no differences between the models. However, unlike the MCAR and MAR models, the MNAR models detected a significant variance in slopes.

6.4 Discussion

Invariably, longitudinal studies have missing data; some respondents drop-out of the study (attrition), miss a wave (wave non-response), or leave some items unanswered (item non-response). In practice, this missing data is ignored, mostly by applying LD, which assumes MCAR (see Table 6.1, Marsh & Hau, 2007). On the other hand, methodologists recommended that a sensitivity analysis be conducted by estimating models that assume missingness was related to either the study's variables (MAR) or to the value, which would have been observed, had the student provided data (MNAR). However, in educational research, there were few practical examples of such sensitivity analysis, and these studies focused (predominantly) on MNAR selection models (Foster et al., 2004; Xu & Blozis, 2011). By using a non-simulated dataset, this study exemplified the estimation of growth in three learning strategies (memorizing, analysing and lack of regulation) during higher education by using nine missing data techniques, which assumed respectively MCAR, MAR and MNAR (see Table 6.1).

6.4.1 Reporting on the findings from study 1

In study 1, only students who we knew had followed a non-delayed trajectory in higher education were included. This allowed us to investigate whether or not taking missingness due to non-response into account, influenced the results and conclusions. Put differently, can results from students with complete data (using LD) be generalised to all students progressing normally during higher education? Examining the summarized results provided in Table 6.9 indicates that for the memorizing and analysing scale, results did not differ between models assuming MCAR, MAR and MNAR. In other words, using LD, correct results for these two scales were obtained.

For the lack of regulation scale, for which the MCAR assumption had been disproven, results did indicate two differences. First, the intercept variance differed in strength between the models (see Table 6.9: LD<MLaux & MIaux), being estimated larger in the MLaux and MIaux models than in the LD model. More importantly, models assuming different missing data mechanisms differed on the significance slope variance (see Table 6.9: Significant: MLaux, MIaux, MNAR models; Not significant: LD, ML and MI). The MIaux and MNAR models indicated that students vary in their change in lack of regulation over time, whilst the LD, ML and MI models suggested that students followed a comparable growth trajectory over time. Thus, for the lack of regulation scale, how non-response is modelled does lead to substantively different conclusions.

Table 6.9: Summary of results from sensitivity analysis

	Intercept	Slope	Intercept variance	Slope variance
Study 1				
Memorizing	=	=	=	=
Lack of regulation	=	=	LD<MLaux & MIaux	Significant: MLaux, MIaux, MNAR models Not significant: LD, ML and MI
Analysing	=	=	=	=
Study 2				
Memorizing	=	Significant: LD & MAR Not significant: MNAR	=	=
Lack of regulation	LD< MAR & MNAR	=	=	Significant: MI, MLaux & MIaux (ML at the verge) Not significant: LD & MNAR
Analysing	LD>MAR & MNAR	Significant: ML Not significant: other models	=	Significant: MNAR Not significant: LD & MAR

Note: "=" signifies that there were no differences between the results of the different models; "LD<MLaux" means that the estimate is larger for the MLaux model than the LD model; "Significant: MI" indicates that the estimate is significant for the MI model; "Not significant: LD" means that in the LD model, the estimate did not result significant

On an aside, we note that for the lack of regulation scale, the results of the models assuming MAR (ML, MI, MLaux and MIaux) did not confirm one another; the slope variance was significant only for the MLaux and MIaux. This finding confirmed that including auxiliary variables can yield different parameter estimates and/or standard errors (Collins et al., 2001). However, this finding was at odds with literature suggesting the correlations between the auxiliary variable and the variable containing missing data, needed to be quite high to matter (81%, Collins et al., 2001; 49%-81%, Johnson & Young, 2011). In our data, only 13.5% to 20.9% of the variance in lack of regulation was explained by the auxiliary variables (see Table 6.5). Nevertheless, given that the MIaux estimates were in line with MNAR models, the auxiliary variables may have reduced bias (Collins et al., 2001) and rendered missingness more ignorable (Allison, 2009; Yoo, 2009). Given that simulation studies did not find harm in including auxiliary variables, which were little related (or unrelated) to the variables under study (Collins et al., 2001), the finding suggests that auxiliary variables should be included when they are available.

How can the findings of the sensitivity analysis for the three scales of study 1 be reported? Table 6.10 summarizes some guidelines that the literature presented (Enders, 2010; Foster et al., 2004; Graham & Schafer, 1999; Jeličić, Phelps, & Lerner, 2010; Muthén et al., 2011).

For the analysing and memorizing scales, showing no differences in average or differential growth between the models (see Table 6.10, option 1), we suggest reporting the results from models assuming MAR. Next, it is worth noting that models, assuming MNAR, did not contradict these findings. For the lack of regulation scale, the results from the MLaux and MIaux model differed from those of the ML and MI models. Moreover, the results from the MIaux model were in line with those of models assuming MNAR (see Table 6.10, option 3). Consequently, it is recommended to present all models assuming MAR and MNAR. Next, a researcher has to choose the model based upon the assumptions with which he/she is most comfortable (Enders, 2010; Foster et al., 2004; Muthén et al., 2011). Here, we would opt to follow the MNAR models and those including auxiliary variables, and thus conclude that students do vary in their growth over time with regard to lack of regulation.

Table 6.10: Guidelines for reporting the results from sensitivity analysis

Result	How to report?
Models assuming MAR ≈ Models assuming MNAR	MAR models in detail Add: Not contradicted by MNAR
Models assuming MAR ≠ Models assuming MNAR	Present MAR and MNAR models Cautiously choose
ML & MI ≠ MLaux, MIaux & MNAR	Present MAR & MNAR models Cautiously choose
ML ≠ MI, MLaux, MIaux & MNAR	Present MAR & MNAR models Opt for the MI, MLaux, MIaux & MNAR results

6.4.2 *Reporting on the findings from study 2*

Study 2 investigated whether taking missingness, due to both attrition and non-response into account, influences the results and the conclusions reached. Put differently, can one, relying on a listwise deleted sample, generalize to all students in higher education? The results of study 2 are summarized in Table 6.9. For the memorizing scale, independent samples *t*-tests did not indicate differences between students from the longitudinal group and their peers with incomplete data. Nonetheless, the models suggested substantively different results; the LD model and the MAR models indicated a significant decline in memorizing over time, while the models assuming MNAR did not confirm this.

To report these results (Table 6.10, option 2), it is recommended to present both the MAR and MNAR estimates. Methodological literature suggests choosing cautiously. Here, we would opt to refrain from stating that students decrease their degree of memorizing. Rather we conclude that students continuing in higher education do reduce their reliance on memorizing strategies, whilst those dropping out after the second wave, retain a constant degree of memorizing.

Concerning the lack of regulation scale, two parameter estimates differed between the models. Firstly, LD underestimated the intercept; this was in line with the rejection of the MCAR assumption for this scale at the first wave. Secondly, the LD and models assuming MNAR did not detect significant slope variance, whilst the models assuming MAR did (see Table 6.10, option 2). In this case we would present both models assuming MAR and MNAR, and we would refrain from concluding that there was slope variance.

Lastly, for the analysing scale, three differences were noted. First, given that MCAR assumption was disproven at the first wave, LD overestimated the intercept. Second, the ML results suggested a significant slope whilst those from MI, MIaux and MLaux did not (see Table 6.10, option 4). Although Jeličić at al. (2010) reported on a comparable finding, the difference between the ML and MI estimates is disconcerting. If the set of cases and the used variables is the same, and if the number of imputed datasets is sufficiently large (here, *m*=100), the ML and MI models should produce equivalent parameter estimates (Collins et al., 2001; Peugh & Enders, 2004). Yet, increasing the number of imputed datasets to 2,000 did not annul the difference between the ML and MI results.

One possible explanation was a violation of the multivariate normal distribution to which ML was found to be more sensitive than MI (Graham & Schafer, 1999; Jeličić et al., 2010). However, there is a lack of guidelines as to whether, in this case, ML or MI is to be trusted more (Jeličić et al., 2010). Therefore, a practical suggestion could be to estimate all four MAR models when auxiliary variables are available and both the ML and MI models when they are unavailable. Reporting on these findings (see Table 6.10, option 4), results from models assuming MAR or MNAR are to be presented. Those from the MI, MLaux, MIaux and MNAR models seem here more plausible given that they confirmed one another.

A third difference in the results of the analysing scale concerns the slope variance. The MNAR models indicated differential growth, while the MAR models did not (see Table 6.10, option 2). Again, the estimates of both models should be reported. Here, we would, from a conservative point of view, opt for the results of the models assuming MAR (no slope variance), to avoid type I error (i.e., stating there is differential growth over time, while in reality there is not).

6.4.3 Implications for analysis of longitudinal data

Two implications for analysis of longitudinal data overarched the two studies. Firstly, the various models assuming different missing data mechanisms (MCAR, MAR and MNAR) led to substantively different conclusions. This was true for both study 1 (missingness due to only non-response, generalizing to the persisting students) and 2 (missingness due to attrition and non-response, generalizing to all students), though it was more outspoken for the latter. For practice, this implies that sensitivity analysis proved valuable for both studies: whenever there is missing data, either due to non-response or due to attrition, estimating multiple missing data models - assuming MAR and MNAR – is recommended.

Second, the LD approach often generated different estimates than the models assuming MAR or MNAR. When the MCAR assumption was disproved for the first wave, the intercept was either over- or underestimated. Next to this, and in line with prior findings (Wothke, 2000; Xu & Blozis, 2011), for some learning strategy scales, the intercept and slope variance was underestimated. As shown repeatedly in simulation studies (Enders, 2001; Enders & Bandalos, 2001; Wothke, 2000), the results from the LD missing data technique were "inadequate at best, misleading at worst" (Jeličić et al., 2010, p. 819). LD should therefore be refrained from, in favour of models assuming MAR (Marsh & Hau, 2007).

6.4.4 Limitations

It has to be acknowledged that there were a number of limitations of this study. Firstly, item non-response was treated as wave non-response, given that there was no easy solution for the item non-response issue (Jeličić et al., 2010). There is a call for further methodological research on dealing with item non-response (Johnson & Young, 2011). Although the amount of missing data, due to item non-response, was very limited (e.g., 6 of 1,031 students were missing for the analysing scale in the first wave, see Table 6.3), excluding cases with item non-response may have decreased statistical power. However, this influence is known to be more substantial when there are a large number of items in a scale (Enders, 2010), which was not the case in our study (all three scales were based upon 4 items, see Table 6.2).

Secondly, the results from this study cannot be generalized to other longitudinal studies or, specifically, to studies on the change in learning strategies during higher education. The issue of missing data is characteristic to each study and to each dataset. Consequently, each study warrants sensitivity analysis to assess the effect of missing data, related to the outcome under study, on the longitudinal change.

Thirdly, administrative data provided us with information on the registration of students in each of three academic years. This allowed us to discern the students who progressed in a non-delayed manner (study 1). Moreover, in study 2, *true* drop-out patterns could be relied upon. When this data is not available, researchers need to construct plausible drop-out patterns on the observed data. However, as missing data is a mixture of non-response and attrition, it can be difficult to discern if and when a student has dropped out.

Notwithstanding this study's constraints, we hope to have provided a clear and practical case of applying sensitivity analysis when missing data is part of modelling change over time. It is apparent from the results that the choice of missing data technique influences the results and the substantive conclusions achieved. This underscores the need to conduct sensitivity analysis when missing data may be related to the outcome under study.

6.5 Mplus 6.1 syntaxes for the memorizing scale in study 2

based on Enders (2010)

For each model, the syntax is provided along with explanation. This explanation is preceded by the "!" symbol, since Mplus does not read lines which start with "!".

6.5.1 MCAR assumption

A. Growth model using LD

! The listwise deleted sample is selected beforehand and converted to a csv-file

DATA: file =
C:\Users\Desktop\Missingness\Memorizing\MemorizingLDsample.csv;
! description of the variables in the dataset, in this case the students' id !number and their score at
! the three waves
VARIABLE:
names = id y1 y2 y3;
! indicate which variables you will use in the analysis
usevariables = y1 y2 y3;
! the actual analysis model, here a growth model with unequal time !intervals (here, 14 months between wave 1 and 2 and 12 months between !wave 2 and 3)

```
MODEL:
icept linear | y1@0 y2@1.16 y3@2.16;
! ask for the TECH4 output which you can consult if there are problems with the model
OUTPUT: TECH4;
! Note: you do not need to define that the analysis needs to use maximum likelihood as it !is the default in Mplus for this analysis
```

6.5.2 MAR assumption

B. Maximum Likelihood

```
! The differences with the growth model on a listwise deleted sample (see A.) are that the !dataset now contains respondents with missing values, !and that the syntax contains a MISSING statement.
DATA: file =
C:\Users\Desktop\Missingness\Memorizing\MemorizingALLsample.csv;
VARIABLE:
names = id y1 y2 y3;
usevariables = y1 y2 y3;
! indicate how Mplus can detect whether data are missing, you can define !the value yourself (see 6.6)
missing = all (999);
MODEL:
icept linear | y1@0 y2@1.16 y3@2.16;
OUTPUT: TECH4;
```

C. Multiple imputation

STEP 1: IMPUTATION

```
DATA: file =
C:\Users\Desktop\Missingness\Memorizing\MemorizingALLsample.csv;
VARIABLE:
names =  id y1 y2 y3;
usevariables = y1 y2 y3;
missing = ALL(999);
```

```
! in the Data imputation command you state that variables y1-y3 need to be imputed and !that you require 100 imputed datasets
DATA IMPUTATION:
impute = y1-y3 ;
ndatasets = 100;
save = memoimp*.dat;
ANALYSIS:     TYPE = BASIC;
OUTPUT:     TECH8;
```

STEP 2: ANALYSIS & POOLING

```
! The memoimplist.dat file, made in step 1, is now used for the analysis
DATA:  FILE = C:\Users\
Desktop\Missingness\Memorizing\memoimplist.dat;
TYPE = IMPUTATION;
VARIABLE:
names =  y1 y2 y3 ;
usevariables = y1 y2 y3;
ANALYSIS:     ESTIMATOR = ML;
MODEL:     i s | y1@0 y2@1.16 y3@2.16;
OUTPUT:     TECH1 TECH4;
```

D. Maximum likelihood with auxiliary variables

```
! For this model, you need the auxiliary variables in your csv-file
DATA: file =
C:\Users\Desktop\Missingness\Memorizing\MemorizingALLsamplewithauxiliary.csv;
VARIABLE:
names are id y1 y2 y3 aux1 aux2 aux3 aux4 aux5 aux6 aux7 aux8 aux9 aux10 aux11 aux12 aux13 aux14 aux15;
usevariables are y1 y2 y3;
auxiliary = (m) aux1 aux2 aux3 aux4 aux5 aux6 aux7 aux8 aux9 aux10 aux11 aux12 aux13 aux14 aux15;
missing are all (999);
MODEL: icept linear | y1@0 y2@1.16 y3@2.16;
OUTPUT: TECH4;
```

E. Multiple imputation with auxiliary variables

STEP 1: IMPUTATION

```
! For this model, you need the auxiliary variables in your csv-file
DATA: file =
C:\Users\Desktop\Missingness\Memorizing\MemorizingALLsamplewithauxil
iary.csv;
VARIABLE:
NAMES =  id y1 y2 y3 aux1 aux2 aux3 aux4 aux5 aux6 aux7 aux8 aux9 aux10
aux11 aux12 aux13 aux14 aux15;
usevariables are y1 y2 y3;
MISSING = ALL(999);
AUXILIARY= aux1 aux2 aux3 aux4 aux5 aux6 aux7 aux8 aux9 aux10 aux11
aux12 aux13 aux14 aux15;
DATA IMPUTATION:
impute = y1-y3 ;
ndatasets = 100
save = memoimp*.dat;
ANALYSIS: TYPE = BASIC;
OUTPUT: TECH8;
```

STEP 2: ANALYSIS & POOLING

The same as in C. Multiple imputation

6.5.3 MNAR assumption

F. Hedeker & Gibbons

```
! For this model, you need a dummy in your csv file discerning 2 groups. Here,
the !dropout variable indicates whether a student progressed normally
throughout their 3 !years of study (0, completers) or not (1, dropout, see 6.2.4.
Plan of Analysis, last paragraph)

DATA: file =
```

```
C:\Users\Desktop\Missingness\Memorizing\MemorizingALLsamplewithdrop
out.csv;
VARIABLE:
names are id y1 y2 y3 dropout;
! Mplus makes a copy of the dropout variable
usevariables are y1 y2 y3 dropout dropout2;
missing are all (999);
nominal are dropout2;
define:  dropout2 = dropout;
ANALYSIS:
estimator = ml;
MODEL:  model:
icept slope | y1@0 y2@1.16 y3@2.16;
y1-y3 (1);
[icept] (b00);
[slope] (b10);
icept on dropout (b02);
slope on dropout (b12);
[dropout2#1] (logit);
MODEL CONSTRAINT:
! first, you define new parameters
new(pic pid iceptc slopec iceptd sloped iceptavg slopeavg );
! computing pattern proportions
pic = exp(logit)/(exp(0) + exp(logit));
pid = exp(0)/(exp(0) + exp(logit));
! the estimates for the completers
iceptc = b00;
slopec = b10;
! the estimates for the dropouts
iceptd = b00 + b02;
sloped = b10 + b12;
! calculating the average estimates across patterns, it are these average estimates
that you report
iceptavg = pic*iceptc + pid*iceptd;
slopeavg = pic*slopec + pid*sloped;
OUTPUT:  sampstat;
```

G. Model with complete case restriction

```
! For this model, you need a variable in your csv-file discerning 3 groups (see 6.2.4. Plan of !Analysis, last paragraph). Here, the pattern variable indicates whether a student was in a !non-delayed trajectory (pattern 1), registered up to the second year (pattern 2, dropout !after wave 2) or registered only in the first year (pattern 3, dropout after wave 1)
DATA: file =
C:\Users\Desktop\Missingness\Memorizing\MemorizingALLsamplewithpatt
ern.csv;
VARIABLE:
names are id y1 y2 y3 pattern;
usevariables are y1 y2 y3 ;
missing are all (999);
! missing data patterns are defined as three known classes
classes = patt(3);
knownclass = patt(pattern = 1 pattern = 2 pattern = 3);
ANALYSIS:
! given you have different classes, you need mixture modeling
type = mixture;

MODEL:
icept slope | y1@0 y2@1.16 y3@2.16;
y1-y3 (1);
[patt#1] (p1logit);
[patt#2] (p2logit);

! models per pattern
! pattern 1
[icept] (p1i);
[slope] (p1s);
! pattern 2
[icept] (p2i);
[slope] (p2s);
! pattern 3
[icept] (p3i);
[slope] (p3s);
```

```
MODEL CONSTRAINT:
! restraining the slope for the group with only 1 datapoint (pattern 3) to the slope of those !in a non- delayed trajectory
p3s = p1s;

! calculating the iceptavg and slopeavg, which are reported
new(c1prop c2prop c3prop iceptavg slopeavg );
c1prop = exp(p1logit)/(exp(0) + exp(p1logit) + exp(p2logit));
c2prop = exp(p2logit)/(exp(0) + exp(p1logit) + exp(p2logit));
c3prop = exp(0)/(exp(0) + exp(p1logit) + exp(p2logit));
iceptavg = c1prop*p1i + c2prop*p2i + c3prop*p3i;
slopeavg = c1prop*p1s + c2prop*p2s + c3prop*p3s;

OUTPUT:  sampstat;
```

H. Model with neighbouring case restriction

```
! compared to model G. complete case restriction, only the restriction in the model constraint !section differs
MODEL CONSTRAINT:
! restraining the slope for the group with only 1 datapoint (pattern 3) to the slope of the !students who are registered up to the second year !(pattern 2)
p3s = p2s;
```

I. Model with available case restriction

```
! compared to model G. and H., only the restriction in the model !constraint section differs
MODEL CONSTRAINT:
! restraining the slope for the group with only 1 datapoint (pattern 3) to the weighted !average of the slopes for pattern 1 and 2. To do so, you need the number of students !showing pattern 1 and 2 (here, respectively 395 and 184)
p3s = (395/(395+184))*p1s + (184/(395+184))*p2s ;
```

7. Conclusion and discussion

Preparing students for lifelong learning is an important challenge for education. To be able to continue learning after graduation, students require a set of skills for lifelong learning. Two frequently mentioned skills are critical thinking and self-regulation. Given that deep, self-regulated learning strategies are required for the development of these skills, education should aim to foster these learning strategies.

Studies investigating whether education meets this aim are however characterized by two main shortcomings. First, studies conducted up to the present are all in the context of higher education, thereby possibly clouding our view of how learning strategies change over time. Longitudinal studies focusing on transitional phases, such as the transition from secondary to higher education, can help us to understand whether growth in learning strategies is an unstable rather than gradual trend. In addition to examining average growth, a limited number of studies have looked at differential growth and detected, for some scales, decreasing variability between students over time. The question of whether trend holds true when differential growth is assessed during a period of change in an educational context remains.

Second, the statistical choices made in current research may compromise the accuracy of estimates and possibly the validity of the conclusions drawn from them. The first issue is the neglect of the measurement model underlying scale scores. This means that measurement error remains unaccounted for. A second element is that, given longitudinal measurement invariance is untested for, change over time may be confounded with measurement variance. A third element concerns omission of respondents with missing data prior to analysis. This can hamper the accuracy of results due to a reduction in statistical power as well as due to selection bias in the remaining sample of respondents with complete data.

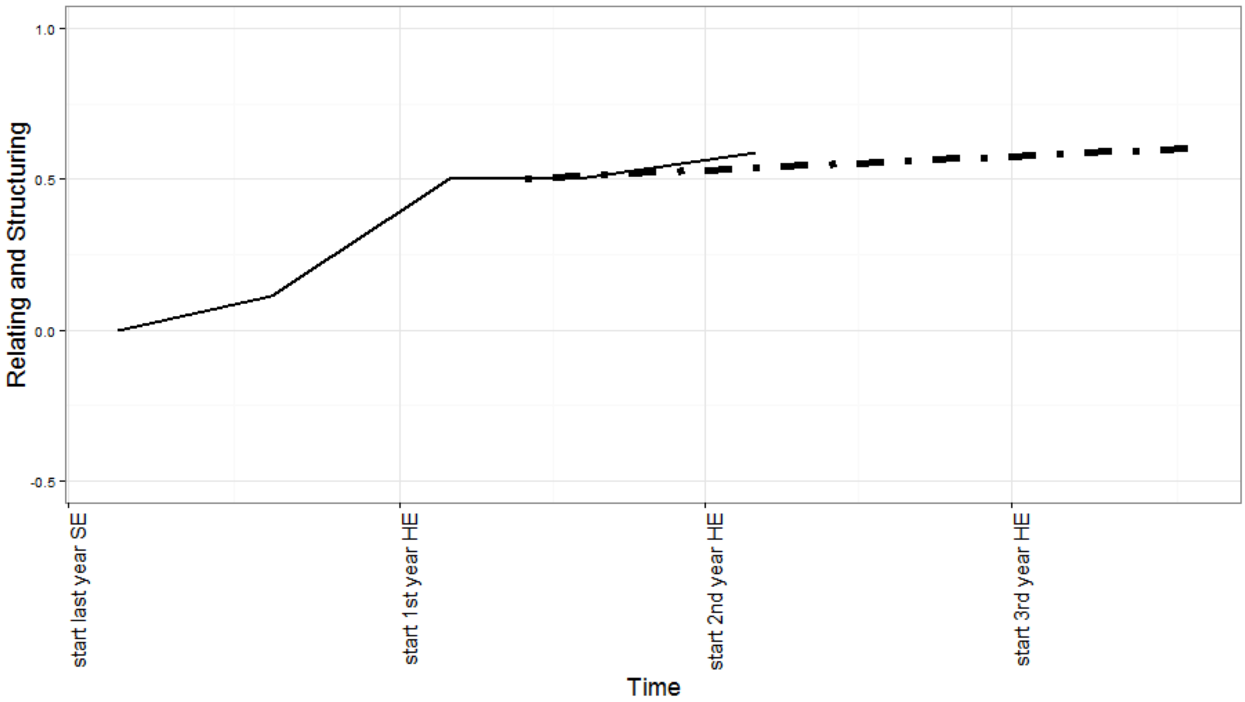
For each of these three statistical choices, statistical literature provides sound alternatives. Use of these alternatives in the students' approaches to learning (SAL) domain is however hampered by an absence of studies detailing whether and how growth trend estimates may be influenced by accounting for the measurement model, by testing for longitudinal measurement invariance and by including missing data into the analysis. Therefore, the present dissertation aims to model the growth in learning strategies during and after the transition to higher education and to explore how statistical choices impact the growth trend estimates.

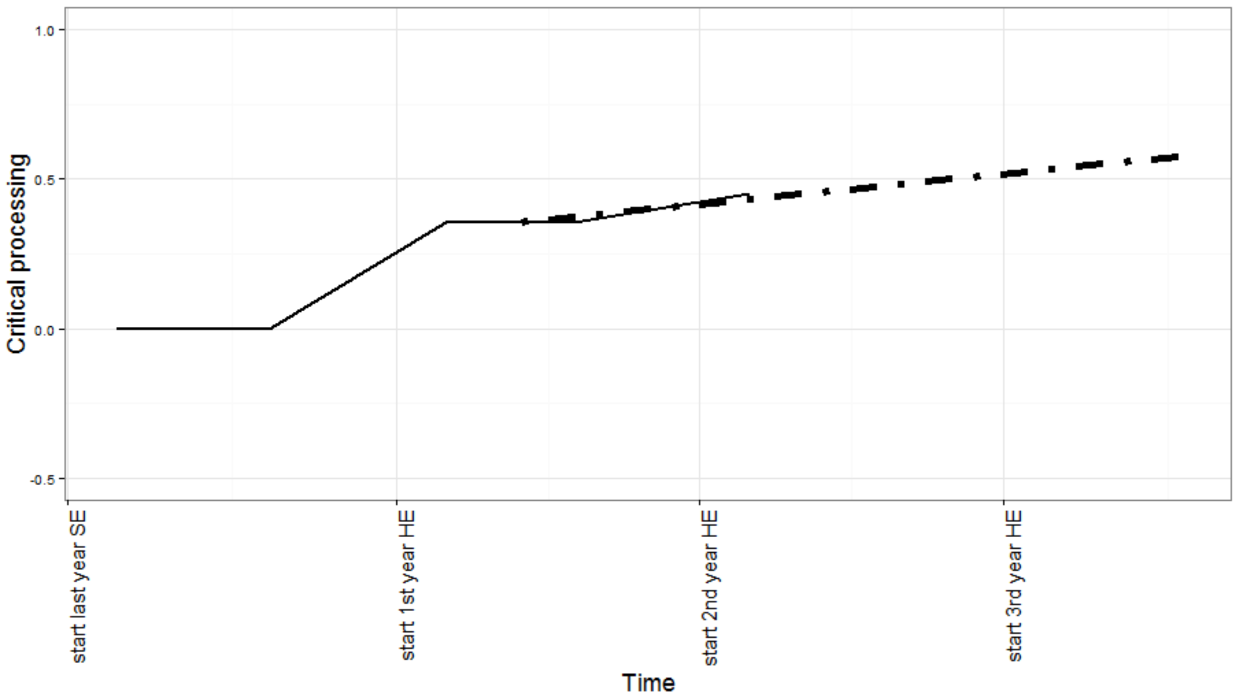
In doing so, seven learning strategy scales of the Inventory of Learning Styles – Short Version (ILS-SV) are used (see 1.1.2). Four of these scales map cognitive processing strategies. The scales used to assess critical processing and, relating and structuring are measuring deep processing, while those assessing memorizing and analysing map stepwise processing. The other three scales - self-regulation, external regulation, and lack of regulation - capture how student learning is guided.

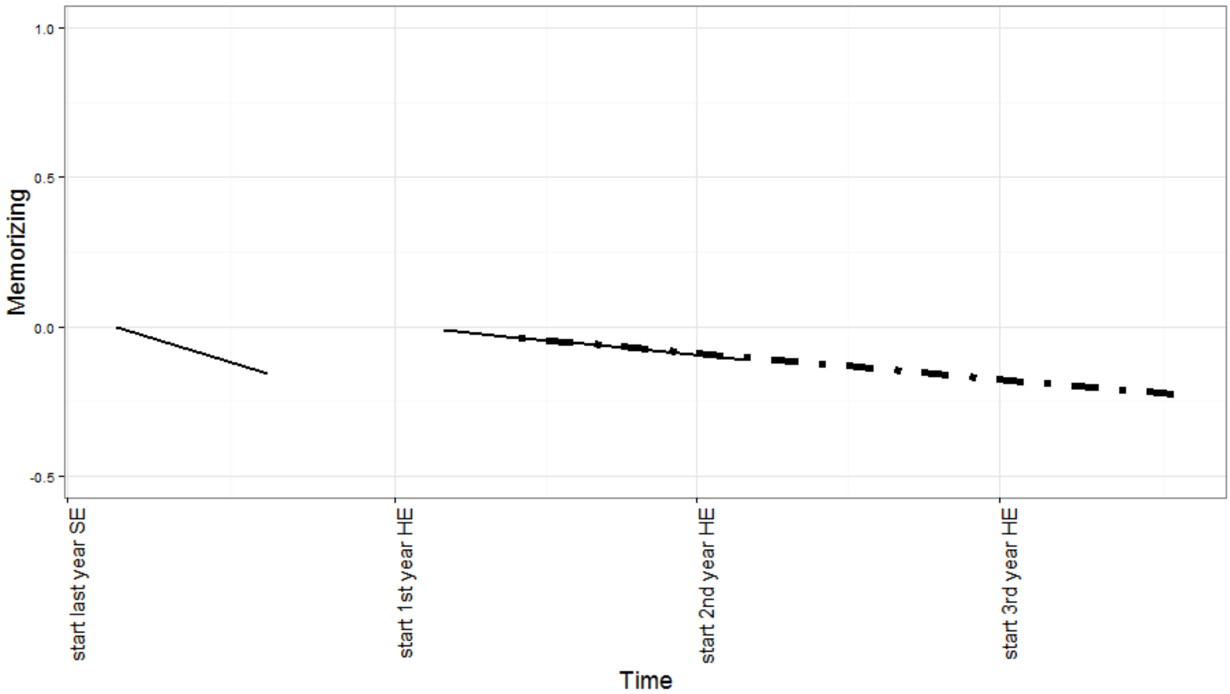
In the next sections, I will first discuss the results of each of the five research questions. In each instance, directions for future research will be linked to the findings. Afterwards, limitations and more general and broad pathways for further research are detailed. Finally, the implications of the present work for research practice on the one hand and for policy and practice on the other hand are discussed.

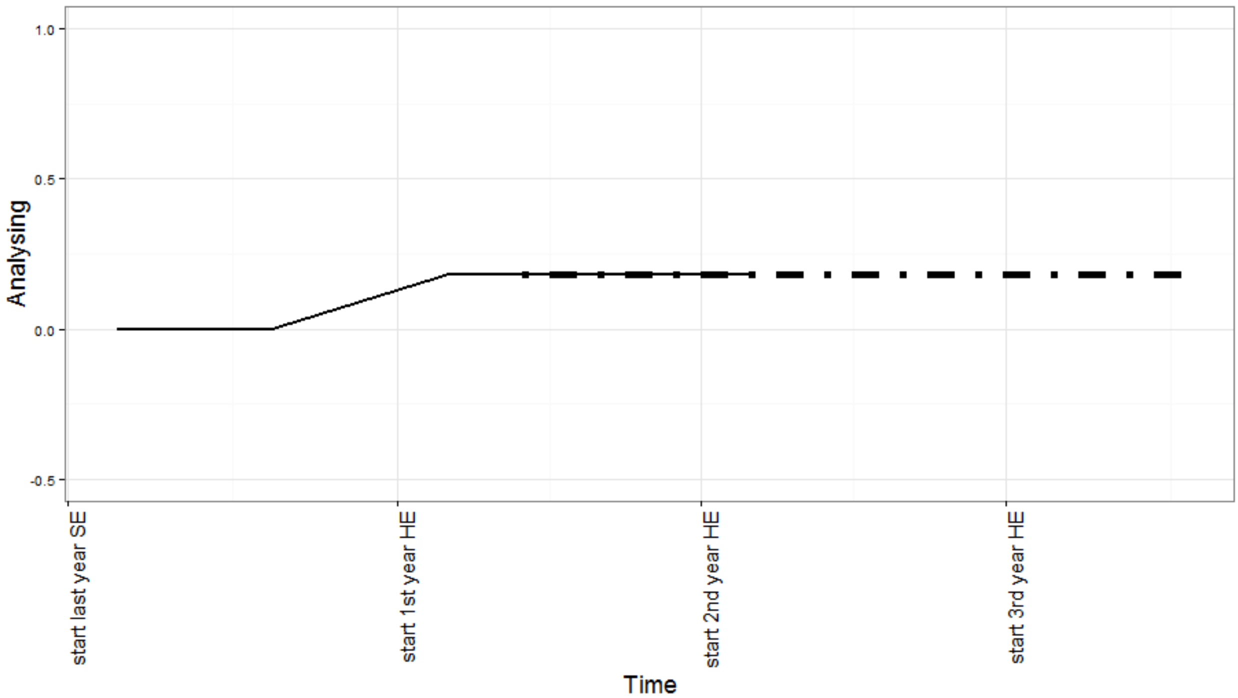
7.1. Average change in learning strategies during and after the transition to higher education

Two studies in this dissertation focused on how students' learning strategies evolve on average during the transition from secondary to higher education and throughout higher education. Chapter four describes the results in a second longitudinal sample, in which one cohort of students was followed up over their three years at a University College. The analysis was performed on a sample of 245 students who provided complete data during each of the three measurement waves. Chapter five explores growth in the learning strategies of students making the transition from secondary to higher education (sample 1). Growth analysis was performed on the 630 students who declared to be studying in higher education during the 18 months following graduation from secondary education. For the external regulation scale in this sample, growth could not be modelled due to insufficient reliability. For the other learning strategy scales, Figure 7.1 visualizes the average growth trajectory from sample 1 (continuous line) and sample 2 (broken line).


1.0
0.5
0.0
-0.5
Relating and Structuring
start last year SE
start 1st year HE
start 2nd year HE
start 3rd year HE
Time


1.0
0.5
0.0
-0.5
Critical processing
start last year SE
start 1st year HE
start 2nd year HE
start 3rd year HE
Time


1.0
0.5
0.0
-0.5
Memorizing
start last year SE
start 1st year HE
start 2nd year HE
start 3rd year HE
Time


1.0
0.5
0.0
-0.5
Analysing
start last year SE
start 1st year HE
start 2nd year HE
start 3rd year HE
Time

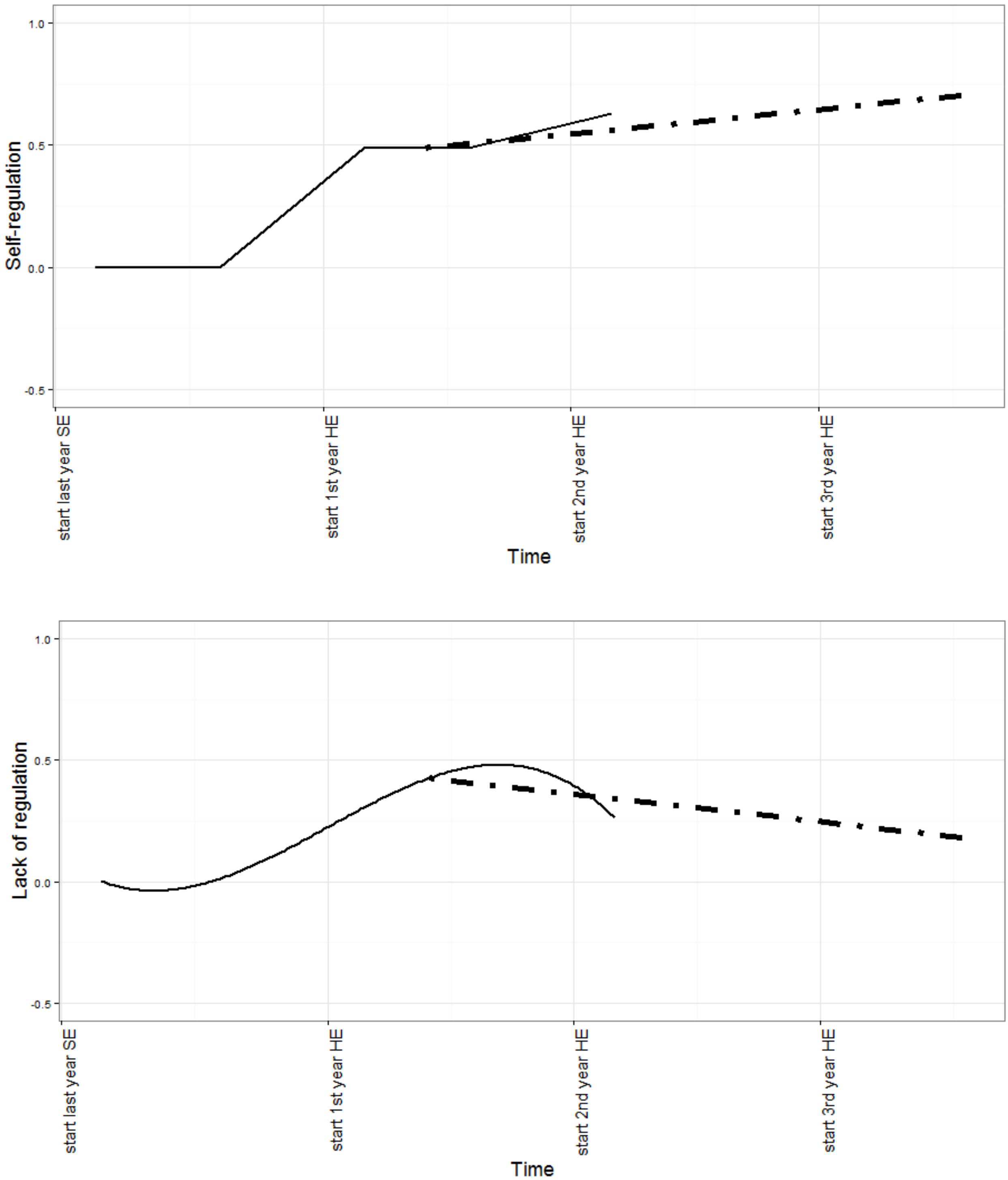


Figure 7.1: Average growth per learning strategy for sample 1 (continuous line) and sample 2 (broken line)

(SE=Secondary education; HE=Higher education)

When the degree of processing and regulation from the start of the last year of secondary education is compared to the end of the third year of higher education, students' learning strategies appear, on average, to evolve in the direction of deep and self-regulated learning. From the viewpoint of lifelong learning in which self-regulatory skills and strategies for deep understanding are pivotal, this is a positive finding: during the period from the last year of secondary education to the end of higher education students appear to become more proficient in these crucial aspects of lifelong learning.

For memorizing, there is a decrease when the level at the start of the last year of secondary education is compared to the end of the third year. During higher education, a decrease in externally regulated learning was also found. Given that these learning strategies are judged - at the conceptual level – to be less beneficial for lifelong learning (see Table 1.2., Vermunt & Vermetten, 2004), this is a positive finding. Similarly, the results from the analysing scale and the lack of regulation scale show that the level at the end of the third year in higher education is higher than during the last year in secondary education. Conceptually, these learning strategies are judged less adequate in regards to lifelong learning, and thus, preferably, these learning strategies had not increased over time.

For both scales, the findings are also at odds with previous research concluding on a constant trend in analysing (Donche et al., 2010; Severiens et al., 2001; Vanthournout, 2011) while being inconclusive between a constant trend (Severiens et al., 2001; Van der Veken et al., 2009) or a decrease (Donche et al., 2010; Vanthournout, 2011) for the lack of regulation scale. This may be due to the timing of the measurement waves. As shown in Figure 7.1, for both scales the increase occurs when students start higher education. If the first data collection takes place when students are already in higher education, as is the case for previous research, this effect can be missed. In sum, the contradictory findings can be explained by the fact that the present research includes the transitional period, while previous research did not.

The findings also show that it does not make sense to view development as a gradual trend. Rather, the results suggest that growth in learning strategies varies across a series of stages. The first stage concerns the last year of secondary education when, for most scales, a constant trend is noted: critical processing as well as self-regulation does not increase and students' degree of lack of regulation also remains constant. Possibly, an explanation can be found in the interplay between teacher-regulation and student-regulation of learning (Vermunt & Verloop, 1999). Perhaps, over time in secondary education, both elements became compatible (i.e., congruence). A consequence of this is that students were not incited to change their learning strategies during the last year of secondary education.

A second stage concerns the transition period from secondary to higher education, when all learning strategies show – compared to the stage before and after - a strong increase, labelled a 'transition jump'. The increase in both deep and stepwise processing strategies can be related to the both-high profile detected by Lindblom-Ylänne and Lonka (1999). These researchers concluded that, in contrast to more advanced peers, novice medical and psychology students most frequently scored high on both meaning- and reproduction-oriented orientation.

In the SAL literature, two contrasting hypotheses have been put forward to explain students' reactions when confronted with a new educational context. The first states that a new educational context makes students rely more strongly on their usual way of going about learning (Cliff, 2000; Segers et al., 2006). The second suggests that a new educational context induces a period of friction, which stimulates students to adjust their learning strategies (Lindblom-Ylänne, 2003; Vermunt & Vermetten, 2004). The transition jump appears to confirm the second hypothesis. Possibly, due to the larger variety of courses compared to in secondary education, the first year in higher education demands more deep and self-regulated learning in addition to greater reliance on analysing and learning some content by heart.

However, if the first year in higher education requires students to rely more on memorizing than in secondary education, we would expect the high level of memorizing to be maintained throughout the first year. This is not the case (see Figure 7.1): after the initial encounter with higher education, memorizing decreases anew. In addition, the friction hypothesis is, in my view, not sufficient to explain the transition jump in the lack of regulation scale. It can hardly be argued that the first year in higher education demands more undirected learning from students.

An alternative explanation may be anxiety, which has previously been found to be associated with student learning (e.g., Tooth, Tonge, & McManus, 1989). Possibly, uncertainty and stress about the learning required in the new educational context, as described in qualitative research (Christie et al., 2008; Cree et al., 2009), may leave students experiencing a greater lack of regulation in their learning. The uncertainty may also have motivated students to maximize their chances by augmenting their use of both deep and stepwise processing activities. In summary, to clarify the transition jump in all processing and regulation strategies, including in lack of regulation, uncertainty and stress generated by the new educational context appears to be a more suitable explanation than the friction hypothesis.

After the relative stability during the last year of secondary education (stage 1) and the transition jump (stage 2), the third stage (the remainder of the first year) is characterised by a relative stand-still. Students maintain their relating and structuring, critical processing and self-regulation strategies constant. Students' lack of regulation continues to increase however, plausibly because until the first year has been passed, uncertainty about whether one can cope with the demands of the new educational context remains. From the perspective of lifelong learning, at the conceptual level, the only positive trend is for the memorizing scale, showing a decrease during this time frame.

In a fourth stage, from the end of the first year of higher education to halfway through the second year of higher education, the change in learning depicts a new change in the direction of deep and self-regulated learning. This appears at the detriment of memorizing and lack of regulation. These trends are continued throughout higher education, as shown by results of the second longitudinal sample, in which students were followed during their three years at a University College (see broken lines in Figure 7.1).

To summarise, from the last year of secondary education to the last year of higher education, on average, students do evolve in the direction of deep and self-regulated learning, to the detriment of memorizing. By extending this study beyond the educational context to include the transitional phase from secondary to higher education, there is evidenced that change in learning strategies is non-linear, with periods of stability (e.g., the last year of secondary education) alternating with periods of change (e.g., the transition jump).

Further research would benefit from exploring why students' learning strategies remain, on average, stable during one phase and change during another. Following students during this transition and throughout the first year of higher education using qualitative methods can reveal the reasons for the transition jump as well as for the relative resistance to change during the rest of the first year in higher education.

Another way to enhance understanding is by studying predictors of growth. Here, literature exploring fit in the teaching/learning environment (T/LE) between secondary and higher education is of value. Research by Torenbeek and colleagues (2010) suggests that students benefit from learning environments not to dissimilar from those that they are already acquainted with in secondary education. If the resemblance in T/LE is high, students need less time to adjust (Torenbeek et al., 2010). Similarly, it has been found that when the T/LE is in line with students' expectations, students engaged more in learning (Stevenson, Sander, & Naylor, 1996). It is worthwhile examining whether the degree of similarity of the T/LE to secondary education or with students' expectations is related to growth in their learning strategies over time, and, more specifically, to the strength of the transition jump.

However, including such predictors for growth is only relevant if students differ in their growth over time (Byrne, 2010). The next section details the findings on differential growth.

7.2. Differential change in learning strategies during and after the transition to higher education

In addition to average growth in learning strategies, differential growth is also of interest: does education incite deep and self-regulated learning in all students? The fifth chapter examines this differential growth during the transition from secondary to higher education; the fourth chapter looks at this element during higher education. The results from both studies are summarized in Table 7.1.

Two overall findings emerge from these results: (1) the absence of differential growth in the second sample, and, (2) a decreasing variability over time for three scales in the first sample. Regarding the former finding, results indicate no differential growth in the second sample across all scales. This finding is at odds with the results from the first sample and contradicts previous findings (Vanthournout, 2011).

Table 7.1: Overview of findings on differential growth

	Sample 1: Transition from secondary education to higher education	Sample 2: higher education
Relating and structuring	Slope variance: significant Covariance: significant and negative	Slope variance: not significant
Critical processing	Slope variance: significant Covariance: significant and negative	Slope variance: not significant
Memorizing	Slope variance: not significant	Slope variance: not significant
Analysing	Slope variance: significant Covariance: significant and negative	Slope variance: not significant
Self-regulation	Slope variance: significant Covariance: not significant	Slope variance: not significant
External regulation	/	Slope variance: not significant
Lack of regulation	Covariance intercept & linear: significant and negative Covariance intercept and quadratic growth parameter: not significant Covariance linear and quadratic parameter: (almost) significant and negative	Slope variance: not significant

A first explanation for this outcome may be that, by mapping the growth of only those students with complete data (listwise deleted sample) there was insufficient statistical power to reject the null hypothesis. Analysing the data in an alternative fashion, allowing for missing data, altered the results for the self-regulation scale only: students were found to vary in their growth over time. For the other 6 scales, conclusions remained unchanged. In summary, removing students with missing data from the analysis does not provide an adequate explanation for the absence of differential growth in the second sample.

A second explanation can be found in the number of waves. The study by Vanthournout (2011) and the first sample included four and five waves respectively, while in the second sample three waves of data are included. As noted by Wu et al. (2010) and by Muthén and Curran (1997), in comparison to a four wave sample, a three wave sample has less power. As a result, if four or more waves of data had been included, differential growth may have been shown.

The second overall finding concerns that, when growth *is* related to students' initial level, students tend to score more alike over time. In the first sample, for the relating and structuring, critical processing, and analysing scales, an average positive growth trend is combined with negative covariance. This implies that during the transition from secondary to higher education, students scoring lower on these scales increase at a faster pace (and vice versa). Put differently, for the deep processing scales, students with poorer initial learning skills in terms of lifelong learning, catch up to their peers. This is remarkable given that students were studying in relatively homogeneous settings in secondary education and subsequently spread out over different study domains in higher education. If these domains affect students' learning strategies, more variation between students rather than less is expected.

The finding of decreasing variability is in line with results from Vanthournout (2011). Broadening the search to longitudinal studies in the SAL domain, negative covariance is also reported (Phan, 2011). Although covariance can be dependent upon the time point chosen for the intercept (Biesanz et al., 2004;

Gottman, 1995), the predominance of negative covariance in studies in the SAL domain may suggest an alternative explanation is required, namely regression to the mean.

Regression to the mean is a statistical artefact caused by random error captured when measuring a concept (Barnett, van der Pols, & Dobson, 2005; Burt & Obradović, 2013). “On any given occasion, high scores will tend to have more positive random error pushing them up, whereas low scores will tend to have more negative random error pushing them down. On the same measure at a later time, (…) the random error is less likely to be so extreme, so the observed score (the same true score plus less extreme random error) will be less extreme” (Shadish et al., 2002, p. 55). For pretest posttest designs with manifest scale scores, the risk of and possible ways to correct for regression to the mean have been extensively described (e.g., Barnett et al., 2005; Rocconi & Ethington, 2009).

However, I am convinced that the negative covariance found here is not due to regression to the mean for two reasons. First, reducing error helps to reduce regression to the mean. One recommendation to reduce error is to use latent variable modelling (Shadish et al., 2002). This recommendation was followed in chapter 5: a multi-indicator latent growth (MILG) model was used, which explicitly modelled measurement error.

Second, to reduce regression to the mean it is recommended to increase the number of measurement waves (Shadish et al., 2002). In the present case, rather than measuring learning strategies at the last year of secondary education and halfway the second year of higher education (i.e., a pretest posttest design), growth was estimated over five waves. As such, regression to the mean would already have occurred from wave 1 to 2. By including an additional three waves, a student's growth over time is therefore assessed more reliably. In summary, it is implausible to use the explanation of regression to the mean to explain the decreasing variability between students over time given the latent variable modeling approach and the inclusion of five waves.

With this, the question then arises on a substantive explanation for this decreasing variability over time in critical processing, relating and structuring and analysing. Regarding the first two scales, plausibly, by choosing the study domain of their interest, students that initially score lower have greater interest in the content of their learning when compared to their secondary education. This can have motivated them towards deeper processing. Student who initially score lower can also experience constructive friction, encouraging them to increase their reliance on deep processing strategies (Vermunt & Verloop, 1999). For students already scoring higher on deep processing, there is congruence between their learning strategies and the demands of the T/LE. Related to this is the finding of a positive association between deep processing strategies and study success during higher education (Richardson et al., 2012). Therefore, students who already rely strongly on these learning strategies are more likely to experience study success and thus be less incited to change their deep processing strategies. As such, differences between students can decrease over time.

Regarding the analysing scale, a substantive explanation is less clear. This component of stepwise processing measures the degree to which students examine the learning material from start to finish and is judged less beneficial from the viewpoint of lifelong learning (see Table 1.2 and Vermunt & Vermetten, 2004). However, certainly when compared to the memorizing strategy of stepwise processing, the inadequacy of this strategy may be called into question. It has also been found predictive of study success (Donche & Van Petegem, 2011; Vermunt, 2005). As such, and in line with the reasoning made for the deep processing scales, students scoring initially lower may experience constructive friction, inciting them to increase their degree of analysing to a greater extent.

The decreasing variability over time is partially contradicted by the lack of regulation scale. The negative covariance between the intercept and the slope parameter indicates that, over the last year of secondary education, differences between students' lack of regulation diminish. Combined with the constant trend during this stage, this is good news: students scoring higher on lack of regulation decrease their reliance to a greater degree. However, in addition to this, the covariance between the linear and the quadratic growth parameter is borderline significant and negative. If interpreted as significant, this implies that students who diminished their reliance on this scale more during the last year of secondary education increase at a faster rate during the first year of higher education. As such, during the actual transition and the first year of higher education, students' scores on the lack of regulation scale become less comparable over time.

Given that lack of regulation has previously been found to be related to drop-out (Vanthournout et al., 2012), this finding gives reason for concern. Students, who made good progress decreasing their lack of regulation during the last year of secondary education, contradict this progress during their transition to higher education. There may possibly be a subgroup of students for whom learning strategies were less crystallized at the end of secondary education, making them more vulnerable to stress and uncertainty due to the change in T/LE when transitioning to higher education.

This can be further examined using growth mixture modelling (Duncan et al., 2006; Jung & Wickrama, 2008), which explores whether there are latent classes in growth over time. If found, class membership can then be used as a predictor for drop-out and academic achievement. This can help the answer the question of which change trajectories lead to a higher probability of drop-out and which growth trends are associated with higher academic achievement.

Moreover, membership of these classes can be predicted by student characteristics (e.g., gender), personality traits and more malleable personal factors (e.g., motivation). At present, there is an extensive research base to show a correlation between these factors and students learning strategies at a given moment (e.g., Donche, De Maeyer, Coertjens, van Daal, & Van Petegem, 2013; Severiens & Ten Dam, 1998; Vermunt & Vermetten, 2004). However, evidence on how these elements influence the *change* in learning strategies over time is, to my knowledge, absent.

In mapping average and differential growth in learning strategies, the measurement model underlying scale scores was taken into account and longitudinal measurement invariance was tested for. The next two sections discuss the impact of these statistical choices on growth trend estimates.

7.3. Do the estimates of the growth trend differ when the measurement model is taken into account?

To determine whether taking the measurement model into account impacts on growth trend estimates, the growth in learning strategies in the first sample was analyzed using repeated-measures ANOVA, multilevel analysis and MILG analysis (see chapter 2). As shown in Table 7.2, all three statistical techniques map average growth, but only the last one takes the measurement model into account. Regarding differential growth, the results from the multilevel analysis can be compared with those from the MILG. The fourth chapter presents a re-analysis of the second sample, which had previously been examined using repeated measures ANOVA (Donche et al., 2010). As such, the differences in average growth can be compared. Three of these scales were re-analysed in chapter six. Due to constraints of more advanced growth models allowing for missing data, manifest scale scores had to be relied upon. The results of the model using listwise deletion (LD) in this sixth chapter can be compared to those from the MILG in chapter four: in both instances, a growth model was used, but the measurement model is only included in the MILG analysis.

Table 7.2: Overview of analyses on the effect of taking measurement variance into account

	Sample 1°			Sample 2		
	Repeated measures ANOVA (Ch.2)	Multilevel (Ch.2)	MILG (Ch.2)	Repeated measures ANOVA (Donche et al., 2010)	MILG (Ch.4)	LG on manifest scale scores (Ch.6)
Average growth	x	x	x	x	x	x
Differential growth		x	x		x	x
Account for measurement model			x		x	
Test for measurement invariance					x	

° first three waves of data

Regarding average growth, the findings from the repeated measures ANOVA, multilevel analysis, MILG and the latent growth (LG) model on manifest scale scores were found to converge regarding the significance of the growth trend. For two out of the seven scales, the multilevel analysis and MILG analysis on sample 1 differed in terms of the strength of this growth. For the external regulation scale, MILG analysis shows a steeper slope than multilevel analysis, while for the lack of regulation scale, the slope is estimated to be shallower. In summary, for average growth and for the two samples under study here, taking the measurement model into account or modelling change in manifest scale scores did not alter the conclusions concerning average growth.

Regarding differential growth, findings for both sample 1 and 2 suggest that for the intercept variance (i.e., the degree to which students vary at the first wave), results converge regarding significance. However, for four scales in sample 1 and for one scale in sample 2, results diverge in relation to strength. Concerning the slope variance, in sample 2, results on slope variance are similar for the MILG and LG model on manifest scale scores. In the first sample however, for four scales, the MILG model detected both a significant slope variance and negative covariance, while the multilevel model did not. In sum, concerning differential growth, results did not differ regarding the significance of effects but did differ in regards to strength for some scales.

One can view the similarities in results between the different analysis techniques as disappointing (e.g., why bother to take the measurement model into account?) or can misinterpret them as a reason for using more traditional statistical techniques (e.g., rely on repeated measures ANOVA given that it was found to lead to the same substantive conclusions as the MILG model). However, as noted by Marsh and Hau (2007, p. 519), "whereas the incorporation of multiple indicators into these latent variable models may not always alter the substantive interpretations, this can only be determined if the multiple indicators are included". In other words, in a different sample, taking the measurement model into account may affect the estimates on growth as well as the substantive conclusions regarding growth. As a result, to avoid biased estimates, it is recommended that the measurement model be accounted while estimating growth over time.

Results also indicate a remarkable difference between estimates from the multilevel model and the MILG: the strength of the slope differs between the two techniques and it appears that MILG is more powerful in detecting differential growth than multilevel models. At first sight, these results appear to contradict the statement that "under a broad set of conditions SEM and MLM longitudinal "growth curve" models are analytically and empirically identical" (Curran, 2003, p. 529). However, the studies underpinning this statement were performed using LG models on manifest scale scores instead of MILG models (e.g., Hox, 2000; Raudenbush, 2001b). To further persuade educational researchers of the benefits of taking the measurement model into account, clearly more simulation studies are needed to clarify under which conditions MILG leads to slopes with differing strength or has relatively more power to detect differential growth (e.g. degree of measurement error and strength of the true change over time).

When the measurement model is taken into account, longitudinal measurement invariance (LMI) can be also tested (Curran, 2003; Stoel & van Den Wittenboer, 2003). The next section explores whether such LMI can affect results regarding change in learning strategies.

7.4. Can measurement variance over time affect the growth trend estimates?

For the two samples used in this dissertation (see 1.4), LMI is examined, verifying the stability in factor loadings as well as in item-difficulty over time. Chapter 3 describes the procedure for LMI testing and illustrates this for the second sample for the change in learning strategies during higher education. The fifth chapter details the results for LMI for the first sample during the transition from secondary to higher education. Note that LMI was not tested for in relation to the external regulation scale in the first sample, due to insufficient reliability during two of the five waves.

Concerning the second sample, for five learning strategy scales, results confirmed complete longitudinal measurement invariance. With regard to the external regulation and analysing scales in sample 2, one and two thresholds, respectively, failed to reveal equivalence over measurement waves. Concerning the first sample, on the transition from secondary to higher education, complete LMI was confirmed for the analysing and critical processing scales. For the memorizing, relating and structuring and lack of regulation scales, the difficulty level of one item varied over measurement moments. For the self-regulation scale, the difficulty level of two items was found to be non-equivalent over time.

For the scales in which complete measurement invariance was detected, a repeated measures ANOVA using manifest scale scores would not be biased by a changing interpretation of items (Steinmetz et al., 2009; Vandenberg & Lance, 2000). Note, however, that the results from repeated measurement ANOVA can still be biased due to neglect of the measurement model (see 7.3) or due to the omission of students with missing data points (see 7.5).

Over the two samples, for six scales, partial measurement invariance was observed: the factor loadings were comparable over time but the difficulty level of one or two items varied over the waves. This can hamper comparison of manifest scale scores, since it is difficult to disentangle genuine changes in the underlying latent variable from nuisance due to shifts in the difficulty level (Burt & Obradović, 2013; Chueng & Rensvold, 2002; Steinmetz et al., 2009). Fortunately, for each of the scales, the minimum of two items with invariant difficulty level was reached (Steinmetz et al., 2009). As such, estimating change over time is still possible but preferably – to improve accuracy - these small variations should be taken into account when modelling growth, for example using MILG (Marsh & Grayson, 1994; Vandenberg & Lance, 2000).

Given that no variances in factor loadings were detected and the variance in item-difficulty proved rather small (i.e. maximum 2 items affected), the "why bother" question surfaces: if LMI did not prove problematic in these two samples, and is unlikely to have altered the substantive conclusions reached, should one bother with testing for it in future samples? The argument given above, that we only know this through verification (Marsh & Hau, 2007), holds true here as well. LMI is specific for each sample (Guttmannova et al., 2008), implying that other samples or populations may show more measurement variance than observed here. In this case, if the measurement invariance assumption is neglected, the validity of conclusions reached from analysis of change may be compromised (Marsh & Hau, 2007; van de Schoot et al., 2012).

In addition to neglecting the measurement model underlying scale scores and ignoring the LMI hypothesis, research on the change in learning strategies is also characterized by exclusion of students with missing data points. The next section will focus to whether this practice can affect growth trend estimates.

7.5. Do growth estimates differ according to the missing data technique adopted?

In the present work, LD was used in chapters 2, 3 and 4. However, in chapter 4, to verify whether the absence of slope variance was due to the listwise deleted sample, the dataset with missing values was re-analysed using maximum likelihood (ML). This technique assumes that data is missing at random (MAR): the probability of missing data is associated with one or several variables in the study. An example is when students who have a higher score on lack of regulation during the first wave also have a higher chance of dropping out of higher education, and are therefore absent in the second wave. The ML technique is used in the fifth chapter as well.

The sixth chapter goes a step further by presenting the range of modern missing-data techniques assuming MAR as well as techniques that assume that data is missing not at random (MNAR). This assumption holds if students who decreased their deep processing from wave one to wave two were also more prone to drop out of higher education prior to wave two. The (unobserved) change over time in deep processing would then predict the chance of missing data. In the sixth chapter, the sensitivity of the estimates to the missing data techniques used, assuming MAR or MNAR, is examined.

A number of conclusions link the findings of these chapters. Regarding average growth, compared to techniques assuming MAR or MNAR, LD was found to provide different estimates for the intercept when the longitudinal sample did not show a random subset of starters (i.e., the assumption of missing completely at random – MCAR - was rejected). Regarding the significance of the slope, there were differences between the MAR and MNAR models. Thus, the choice of missing data technique led to different conclusions.

Concerning differential change, LD was found to give a lower estimate of the intercept variance. Regarding the slope variance, in chapter 4, all seven scales had a non-significant slope variance using LD. However, when missing data were included using ML, for one of the seven scales, the slope variance was significant. In chapter 6, differences were also noted between the missing data techniques for this slope variance: the models assuming MAR and MNAR led to different conclusions.

The fact that different conclusions may be drawn depending on the missing data technique used underlines the need for sensitivity analysis. By estimating multiple missing data models, assuming MAR and MNAR (Enders, 2011; Molenberghs & Fitzmaurice, 2009), it becomes clear where disagreements between models lie, and hence where caution in interpretation and generalization is required. On the other hand, the findings regarding which the models agree can be considered trustworthy.

7.6. Limitations and general directions for further research

This dissertation is subject to a number of constraints. The first concerns the fact that only three waves of data were included in the second sample, thereby only allowing constant and linear growth trajectories to be estimated. As such, discerning stages of faster or slower growth was not possible (Metha et al., 2004; Wu et al., 2010), and so limits our understanding of how learning strategies change during the second and third years in higher education. Moreover, as mentioned in 7.2., this may also have hampered estimation of differential growth over time.

A second limitation concerns the estimation of required sample size (N) or power obtained given a certain sample size. This has not been thoroughly addressed in this dissertation. Questions remain as to what is the minimum sample required to test for longitudinal measurement invariance, to model growth over time or to examine non-linear growth. Also unclear is the power obtained given a certain sample size.

It is clear that these questions do not have a straightforward answer; a researcher seldom finds recommendations regarding their specific case in simulation studies. For example, one can find suggestions on how to adjust cut-offs for fit indices for LG models to the sample size and number of waves (De Roche, 2009), but for MILG model such simulation studies are to my knowledge absent. Other elements which can affect power include the number of items per factor, number of Likert points, the non-normality of the item scores and the amount of missing data.

This has led researchers to state that "because requisite sample size is closely tied to the specific model and data of a given study, general rules of thumb are of limited utility" (Brown, 2006, p. 389). Instead, it is recommended that a Monte Carlo simulation is used to calculate the required N. When data has already been collected, such simulation techniques can be used to determine power (Brown, 2006; Burt & Obradović, 2013; Muthén & Muthén, 2009a). For researchers relying on Likert-type self-report questionnaires to map longitudinal change, an example of such Monte Carlo simulation would without doubt be of interest.

A third limitation that needs to be acknowledged concerns the fact that each learning strategy scale was modelled separately. Given that learning strategies are related to one another (Vermunt & Vermetten, 2004), a number of substantively relevant research questions remain unexplored. For example, is the correlation between two scales invariant over time? Or, are these concepts more strongly associated at a certain stage than at other stages?

Moreover, in the learning pattern model, regulation strategies are viewed as antecedents of processing strategies (Vermunt, 1998). This implies that, for example, students' degree of self-regulation at a given time should predict their change in deep processing. Alternatively, to confirm such a theoretically expected relationship, students' level of deep processing should not predict growth in self-regulation. In summary, by relating growth in learning strategy scales to one another, the theoretical relation put forward by Vermunt (1998) can be tested.

In addition the limitations of the present work, a number of general directions for further research can be offered. These can be grouped into substantive and methodological challenges. Regarding the former, more insight into links between learning strategies during higher education and lifelong learning is demanded. In the present work, the average growth trend is toward deep and self-regulated learning. According to Vermunt and Vermetten (2004), at the conceptual level (see Table 1.2.), a stronger reliance on deep and self-regulated learning is beneficial in terms of lifelong learning. Empirical studies have shown that students' habitual ways of learning in higher education are predictive of subsequent learning in the workplace (McManus, Keeling, & Paice, 2004) and that a deep learning strategy is beneficial for career success (Hoeksema et al., 1997). It is however clear that more empirical work is needed to evidence this claim.

Conceptually, memorizing, analysing, external regulation and lack of regulation have all previously been judged as less beneficial for lifelong learning (Vermunt & Vermetten, 2004). There is also evidence that rote learning hampers career success (Hoeksema et al., 1997). For the other learning strategies, empirical proof of added value for lifelong learning appears lacking. As detailed in section 7.2., in my view and possibly depending on the context, the analysing learning strategy may be beneficial for lifelong learning. Aside from the context, the sequence in learning strategies may also be relevant. For example, does the added value of rote learning at the workplace increase when it is followed by deep learning? In summary, the SAL field would strongly benefit from further empirical evidence exploring the link between students' learning strategies during higher education and lifelong learning.

To do so, longitudinal research is required that follows students during secondary education, higher education and during a part of their working life. Such datasets would also allow exploration of the transitional phase from (higher) education to working. To do so in an in-depth fashion, a multi-method design seems most appropriate. Interviews exploring change in learning strategies can provide a complementary picture to quantitative data.

Longitudinal study from secondary education through to working life would require an extensive time frame (e.g., McManus et al., 2004). Using a sequential cohort or accelerated design, in which different groups are followed up over overlapping time frames, can make such data gathering practical, whilst also allowing for growth modelling (Duncan et al., 2006; McArdle & Nesselroade, 2013; Muthén & Muthén, 2010; Tomarken & Waller, 2005). A pitfall in this research can be that no one questionnaire is apt for all contexts (e.g., a learning strategy questionnaire focusing on a higher education setting versus a work setting). Consulting the methodological literature may provide a solution: by allowing for overlap of some items over two consecutive waves while varying others, learning strategies can be modelled over changing contexts (Muthén & Muthén, 2010).

Regarding methodological directions for future research, two important points can be discerned. First, measures of effect size are lacking in relation to LG models (Richardson, 2013). We can discern whether students change their reliance on a certain learning strategy on average and we can compare this growth rate over learning strategy scales, but cannot yet determine how meaningful this change is. For example, in section 7.1., we noted that for a number of scales, the change in learning strategies is stronger during the transition to higher education than afterwards. But, should we view this as a strong increase followed by a small increase or rather as a small one followed by a negligible one? Given the importance of effect sizes for applied research and practice, methodological research on measures of effect size for LG and MILG models is required.

A second point regards the operationalization of learning strategies. The present work focused on how statistical choices can threaten validity when estimating growth in learning strategies using self-report questionnaires. In addition, the predominance of data from self-report questionnaires can constitute a threat to validity as well. When most studies on learning strategies use self-report inventories, this method can affect the results obtained regarding these concepts, which is labelled a mono-method bias (Shadish et al., 2002). For example, respondents can choose the same Likert point but for different reasons (Baxter

Magolda, 1998). Or, over time, there can be qualitative change. Even when the same questionnaire is used over the different waves and longitudinal measurement invariance holds, the meaning given to the constructs can vary over time (Wu et al., 2010).

Here, techniques other than self-report questionnaires can be beneficial to our understanding. Though there have been studies complementing self-report questionnaires with interviews or structured learning reports (e.g., Endedijk & Vermunt, 2013; Schatteman et al., 1997), more are welcomed. Moreover, other methods also appear promising, such as think aloud in which students are asked to verbalize their learning strategies while executing a study task (Ericsson & Simon, 1993). Another is eye tracking followed by cued recall, which consists of recording where a student looks and for how long during a study task. Then, by visualizing the eye movement, the student is prompted to report which learning strategies were used during the task (cued recall) (Azevedo, Moos, Johnson, & Chauncey, 2010). In sum, to avoid a mono-method bias, future research should focus on applying different techniques and integrating findings from each.

7.7. Implications for research practice

A number of implications for research practice can be put forward based on the present work. We note that these implications are not only restricted to researchers in the SAL domain, but are relevant for researchers investigating longitudinal change using self-report Likert-type questionnaires, regardless of the research topic.

A first recommendation is to include the measurement model underlying a scale score when estimating growth over time using a MILG model. As such, the error associated with items is partitioned out from the true score. To account for the correlation between errors over time, these correlations should be modelled explicitly. How this can be done in Mplus is shown in section 3.5. By modelling error and correlated error over time, the estimates of average and differential growth are more accurate. Certainly, when the concept under study has lower reliability levels (e.g. Cronbach's alpha below .80), modelling error at the item level is important when estimating growth (Shadish et al., 2002).

A second recommendation is to test for LMI prior to modeling growth over time, in order not to confound change in measurement with true change. The consequences of longitudinal measurement variance depend on the type and severity of the variance detected. If variance in factor loading is detected, modelling growth should be avoided (Steinmetz et al., 2009). In such instances, one can consider excluding the item in question and re-testing for LMI. When this procedure is not possible due to a limited number of items per scale (e.g., most scales from the ILS-SV have 4 items, which can be considered a minimum), modelling growth over time was not possible due to longitudinal measurement variance.

When the factor loadings prove invariant over time, but item difficulty level varies over time, longitudinal analysis is not always impeded. Here, it is recommended that the difficulty level of at least two items is invariant over time (Steinmetz et al., 2009). If this minimum is reached, estimating change over time is still possible if variations are explicitly modelled in a MILG (Marsh & Grayson, 1994; Vandenberg & Lance, 2000).

What sample size is required to estimate this MILG model? This question does not have a straightforward answer, and, to my knowledge there is no simulation study explicitly addressing this issue. Ideally, as mentioned, the required sample size is estimated through Monte Carlo simulation routines (see 7.6, Muthén & Muthén, 2009a). What does appear from methodological literature are rough ideas on minimum sample size.

Broadening the search to LG models (on manifest scale scores, De Roche, 2009) and to confirmatory factor analysis (Flora & Curran, 2004), it seems that the smallest sample size for which these have been studied is N=100. These studies were also on complete data. Therefore, though including respondents with missing data points can increase power, it is recommended to have at the very least 100 respondents with complete data. However, I expect convergence problems to occur with this sample size. These can make it impossible to assess anything other than very simple models. In sum, to make re-analysis of an existing dataset using MILG worthwhile, 100 respondents with complete data appears a minimum.

However, when planning a longitudinal data analysis, a sample size larger than one hundred should clearly be targeted. Simulation studies frequently include small sample sizes of 200 (Muthén et al., 1997) and 250 (Beauducel & Herzberg, 2006; De Roche, 2009; Nussbeck, Eid, & Lischetzke, 2006). If variables under study are not too skewed, such sample size can be sufficient (Muthén et al., 1997). DiStefano and Hess (2005) detail how the severity of skewness can be determined. Thus, when setting out a longitudinal study, aiming for 200 or more respondents with complete data is recommended (Byrne, 2010). When non-normality is to be expected for Likert-type items, for example based on information from previous data collections, the sample size should be increased to maintain a sufficient level of power (Muthén & Curran, 1997).

Please note that the stated sample sizes refer to when the longitudinal data collection has finished. Longitudinal data collections almost always have missing data. Moreover, the percentage of missing data is likely to increase with the time frame of the longitudinal study. To estimate how much larger a sample needs to be to retain power despite drop out and non-response, one can examine the missing data rates in studies with comparable respondents spanning similar time frame.

A third area in which recommendations regarding statistical methods can be formulated concerns the handling of missing data. A first recommendation is to refrain from LD but rather include respondents with missing data. This can be done using the ML or using multiple imputation (MI; Enders, 2010; Marsh & Hau, 2007). Nowadays, both techniques are available using a number of software packages (e.g. R or Mplus). A researcher can choose either one. In most cases, the results for both techniques will be very similar.

By and large, when making the effort to contact respondents at multiple points over time, we do not restrict ourselves to measuring just one concept. Data on other concepts are usually gathered as well (e.g. gender, predictors for the concept under study or possible consequences of the concept under study – here, for example study success). These variables can be very helpful in estimating the missing data points as auxiliary variables. Given that simulation studies have shown that including auxiliary variables that were minimally related or even unrelated to the variables under study did not bias the results obtained (Collins et al., 2001), it is recommended that auxiliary variables are included in ML or MI whenever they are available.

When it is plausible to assume that missing data is related to the unobserved change in the concept of interest over time (outcome-related missing data, MNAR, e.g., students that decrease their reliance on memorizing have a higher chance of persisting in higher education), it is recommended to estimate multiple missing data models - assuming MAR and MNAR (Enders, 2011; Molenberghs & Fitzmaurice, 2009). Such sensitivity analysis will reveal more solid findings, when the MAR and MNAR models converge, and less solid findings, when estimates differ and lead to substantively different conclusions.

To my knowledge, models assuming MNAR have only been described for manifest scale scores. Taking the measurement model and small measurement variance into account when estimating models assuming MNAR is not possible given current statistical software. For this reason, preferably, scales included in the comparison of MAR and MNAR techniques should show complete measurement invariance over time. If they show partial measurement invariance, a first comparison should be made between models assuming MAR acknowledging the measurement model on the one hand and a model assuming MAR but relying on manifest scale scores on the other hand. As such, the effect of the inequivalence over time on the growth estimates can be gauged prior to assessing the effect of different MAR and MNAR missing data techniques.

A last recommendation concerns the number of data collections. When planning a longitudinal design, I recommend including four or more waves of data. Compared to three waves, this provides more power in detecting differential growth over time (Wu et al., 2010). Moreover, when non-linear growth is expected or possible, this can only be explored using four or more waves of data (Muthén & Muthén, 2010).

When choosing the timing for the data collections, relative stability or variability should be considered. During periods where more change in the construct under study is to be expected, the time interval between waves should preferably be shorter and vice versa (Singer & Willet, 2003). To be able to connect findings on growth, all studies, including those considering linear growth, should detail timing of waves. If not, it remains difficult to discern whether results confirm or rather contradict prior findings when non-linear change is found.

To conclude, we list the eight recommendations put forward:

1. Include the measurement model underlying the scale scores and explicitly model correlations between the errors over time;

2. Test for LMI and only proceed to model growth for a scale if all factor loadings are equivalent over time and the difficulty level for at least two items;
3. If the item difficulty varies over time, model this explicitly when estimating growth over time;
4. Target at least 200 to 250 respondents with complete data at the end of the longitudinal data collection period;
5. Include respondents with missing data using ML or MI instead of LD;
6. Include auxiliary variables in the ML or MI technique whenever available;
7. When outcome-related missing data is plausible, test for the sensitivity of results to missing data technique by estimating models assuming MAR and MNAR;
8. Include four or more waves of data and shorten the time interval between them during periods in which more change is expected.

7.8. Implications for policy and practice

The present work modelled growth in learning strategies during and after transition to higher education and explored how statistical choices can affect growth estimates. From this endeavour, three implications for policy and practice are outlined.

First, the transition jump detected during the change from secondary to higher education has important implications for student guidance. Though, as noted, more in-depth insight into this transition jump is crucial, the period during which students encounter their new educational context of higher education appears ideal for remedial action. Firstly, increased malleability in learning strategies during this period can be used to direct students towards more qualitative and efficient learning strategies. Secondly, students were found to increase learning strategies detrimental to lifelong learning skills as well. The reliance on memorizing and reports on lack of regulation also increase during this stage. This should not be viewed as a temporary upsurge: on average, by halfway through the second year of higher education, the degree of memorizing and lack of regulation is still higher compared to at the end of secondary education. In summary, to take advantage of increased malleability and halt detrimental changes in learning strategies, study skills guidance should be scheduled to tie in with the onset of the first year of higher education.

A second implication for practice regards student drop-out or its inverse, student retention. A recent Eurydice report (European Commission/EACEA/Eurydice, 2014) identified this retention to be a key performance indicator for higher education, given the psychological, financial and emotional costs associated with drop-out. As a result, this report makes a plea for more evidence-based means of targeting of at-risk students. The report notes that "typically guidance and counselling services are too stretched by increased demand to be able to target and reach those most in need" (European Commission/EACEA/Eurydice, 2014, p. 11).

Research indicates that student characteristics at a given moment in time (e.g., upon entry to higher education) are associated with persistence in higher education (e.g., Mäkinen, Olkinuora, & Lonka, 2004; Vanthournout et al., 2012).

The results from the present dissertation indicate that an absence of change in learning strategies can also be predictive of drop-out. For the memorizing scale, students persisting in higher education reduced their reliance on memorizing strategies from the first to the second wave of data collection, whilst those dropping out after the second wave kept their degree of memorizing constant. Comparably, Lonka et al. (2004) reasoned that, in contrast to the first year, third year students required more organizational skills. When students fail to develop these skills from the first to the third year, they will have trouble succeeding.

When attempting to develop evidence-based and targeted guidance services, a thorough analysis of which students have a higher probability of dropping out is required. The present findings suggest to include in this analysis the change over time in malleable student characteristics (such as learning strategies) in addition to student characteristics at a given moment (e.g. start of the first year).

A third implication regards the content of learning strategies guidance. The present work measured growth in students' learning strategies, but it remains unclear whether students themselves were conscious about those changes. In addition, it remains unclear whether students are conscious of desirable learning strategies and their association with drop-out and academic achievement. An important step in guiding students towards adopting learning strategies more beneficial for lifelong learning thus appears to be in raising their awareness by presenting them with results obtained from repeated measurement of their learning strategies over time and contrasting these to results from research on student drop-out and study success.

When setting up coaching initiatives, the importance of taking into account student preferences and expectations has been raised (Sander, Stevenson, King, & Coates, 2000). Concerning learning strategies, initial evidence indeed suggests that students' preferences for feedback on their learning strategies depends on their self-efficacy and regulatory strategies (Donche, Coertjens, Vanthournout, & Van Petegem, 2012). For example, students scoring high on lack of regulation and having a low self-efficacy, tend to prefer external sources for feedback, such as guidance by a mentor or following a course. Conversely, high efficacious and regulated learners, show a stronger preference for self-improvement. In sum, to set up effective guidance on learning strategies, it is recommended to tailor initiatives to students' varying preferences for feedback.

To conclude, I take the liberty of updating the two quotes given at the start of this dissertation.

"An important finding is that the growth in learning strategies during and after the transition to higher education is in the direction of deep and self-regulated learning. But, by including the transitional phase from secondary to higher education, it is evidenced that this is not a one-way, gradual process in which students become more self-regulated, deep-level learners. Rather, it is a capricious pattern, with periods of stability followed by periods of change. Furthermore, it is not a "one growth trend for all"-process: students differ in their developmental patterns over time."

"It is not only unfortunate that the methods by which longitudinal data regarding average and differential change in learning strategies are often analysed are not commensurate with the level of effort involved in their collection. These statistical choices can also affect the growth estimates, thus hampering our understanding of change in learning strategies. As such, it is recommended that the measurement model underlying scale scores is taken into account, longitudinal measurement invariance is tested for prior to modelling growth and that respondents with missing data are included in the analysis."

References

Allison, P. D. (2009). Missing Data. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE Handbook of Quantitative Methods in Psychology* (pp. 72-89). Thousand Oaks, CA: Sage Publications Inc.

Asikainen, H. (2014). *Successful learning and studying in biosciences*. Helsinki: Unigrafia.

Asparouhov, T., & Muthén, B. (2006). Robust Chi square difference testing with mean and variance adjusted test statistics. Mplus Web Notes. Vol. 10. *Mplus Web Notes* (Vol. 10).

Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring Cognitive and Metacognitive Regulatory Processes During Hypermedia Learning: Issues and Challenges. *Educational Psychologist, 45*(4), 210-223. doi: 10.1080/00461520.2010.515934

Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review, 5*(3), 243-260. doi: 10.1016/j.edurev.2010.06.001

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology, 34*(1), 215-220. doi: 10.1093/ije/dyh299

Baxter Magolda, M. (1998). Learning and gender: Complexity and possibility. *Higher Education, 35*(3), 351-355. doi: 10.1023/a:1003166418248

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*(2), 186-203. doi: 10.1207/s15328007sem1302_2

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The Role of Coding Time in Estimating and Interpreting Growth Curve Models. *Psychological Methods, 9*, 30-52. doi: 10.1037/1082-989x.9.1.30

Biggs, J. (1979). Individual differences in study processes and the quality of learning outcomes. *Higher Education, 9*, 114-125. doi: 10.1007/BF01680526

Biggs, J. (1987). *Student approaches to learning and studying*. Melbourne: Australian Council for Educational Research.

Biggs, J. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology, 63*(1), 3-19. doi: 10.1111/j.2044-8279.1993.tb01038.x

Biggs, J. (2001). Enhancing Learning: A Matter of Style or Approach. In R. J. Sternberg & L. Zhang (Eds.), *Perspectives on Thinking, Learning, and Cognitive Styles*. London: Lawrence Erlbaum Associates.

Biggs, J. (2003). *Teaching for Quality Learning at University* (2nd ed.). Buckingham: SRHE & Open University Press.

Biggs, J., Kember, D., & Leung, D. (2001). The revised two-factor Study Process Questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*(1), 133-149. doi: 10.1348/000709901158433

Bijleveld, C., van der Kamp, L., & Mooijaart, A. (1998). *Longitudinal data analysis: Design, models and methods*. London, United Kingdom: Sage Press.

Boekaerts, M., & Martens, R. (2006). Motivated learning: what is it and how can it be enchanced? In L. Verschaffel, F. Dochy, M. Boekaerts & S. Vosniadou (Eds.), *Instructional Psychology: Past, Present and Future Trends* (pp. 113-130). Amsterdam: Elsevier.

Boekaerts, M., Pintrich, P., & Zeidner, M. (2000). *Handbook of Self-regulation* San Diego: Academic Press.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach.* Hoboken, NJ: Wiley.

Boyle, E., Duffy, T., & Dunleavy, K. (2003). Learning styles and academic outcome: The validity and utility of Vermunt's Inventory of Learning Styles in a British higher education setting. *British Journal of Educational Psychology, 73*(2), 267-290. doi: 10.1348/00070990360626976

Breuer, K., & Eugster, B. (2006). Effects of training and assessment in vocational education and training (VET): Reflections on the methodology of assessing the development of traits of self-regulation. *Studies in Educational Evaluation, 33*(3), 243–261. doi: 10.1016/j.stueduc.2006.08.001

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality, 54*(1), 106-148. doi: 10.1111/j.1467-6494.1986.tb00391.x

Brooks, R., & Everett, G. (2008). The impact of higher education on lifelong learning. *International Journal of Lifelong Education, 27*(3), 239-254. doi: 10.1080/02601370802047759

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Bruinsma, M., & Jansen, E. (2009). When will I succeed in my first-year diploma? Survival analysis in Dutch higher education. *Higher Education Research and Development, 28*(1), 99-114. doi: 0.1080/07294360802444396

Burt, K. B., & Obradović, J. (2013). The construct of psychophysiological reactivity: Statistical and psychometric issues. *Developmental Review, 33*(1), 29-57. doi: 10.1016/j.dr.2012.10.002

Busato, V., Prins, F. J., Elshout, J., & Hamaker, C. (1998). Learning styles: A cross-sectional and longitudinal study in higher education. *British Journal of Educational Psychology, 68*(3), 427-441. doi: 10.1111/j.2044-8279.1998.tb01302.x

Byrne, B. M. (2010). *Structural equation modeling with AMOS*. New York, NY: Routledge.

Cheema, J. R. (2014). A Review of Missing Data Handling Methods in Education Research. *Review of Educational Research*. doi: 10.3102/0034654314532697

Christie, H., Tett, L., Cree, V. E., Hounsell, J., & McCune, V. (2008). 'A real rollercoaster of confidence and emotions': Learning to be a university student. *Studies in Higher Education, 33*(5), 567-581. doi: 10.1080/03075070802373040

Chueng, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255. doi: 10.1207/S15328007SEM0902_5

Cliff, A. F. (2000). Dissonance in first-year students' relfections on their learning. *European Journal of Psychology of Education, 15*(1), 49-60. doi: 10.1007/BF03173166

Coertjens, L., Donche, V., De Maeyer, S., Vanthournout, G., & Van Petegem, P. (2012). Longitudinal measurement invariance of learning strategy scales: Are we using the same ruler at each wave? *Journal of Psychoeducational Assessment, 30*(6), 577-587. doi: 10.1177/0734282912438844

Coertjens, L., Donche, V., De Maeyer, S., Vanthournout, G., & Van Petegem, P. (2013a). Modeling change in learning strategies throughout higher education: A multi-indicator latent growth perspective. *PLoS ONE, 8*(7), e67854. doi: 10.1371/journal.pone.0067854

Coertjens, L., van Daal, T., Donche, V., De Maeyer, S., Vanthournout, G., & Van Petegem, P. (2013b). Analysing change in learning strategies over time: A comparison of three statistical techniques. *Studies in Educational Evaluation, 39*, 49-55. doi: 10.1016/j.stueduc.2012.10.006

Coffield, F., Mosley, D., Hall, E., & Ecclestone, K. (2004). *Learning Styles and Pedagogy in post-16 learning. A systematic and critical review.* Witshire, United Kingdom: Learning and Skills Research Centre.

Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd edition ed.). New Jersey, NJ: Lawrence Erlbauw Associates.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cohen, L., Manion, L., & Morrison, K. (2008). *Research methods in education.* (sixth edition ed.). London: Routledge.

Collins, L. M., Schafer, J. L., & Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330-351. doi: 10.1037/1082-989X.6.4.330

Cooke, D. J., Kosson, D. S., & Michie, C. (2001). Psychopathy and ethnicity: Structural, item, and test generalizability of the Psychopathy Checklist - Revised (PCL-R) in Caucasian and African American participants. *Psychological Assessment, 13*(4), 531-542. doi: 10.1037111040-3590.13.4.531

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104. doi: 10.1037/0021-9010.78.1.98

Cree, V., Hounsell, J., Christie, H., McCune, V., & Tett, L. (2009). From further education to higher education: Social work students' experiences of transition to an ancient, research-led university. *Social Work Education, 28*(8), 887-901. doi: 10.1080/02615470902736741

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38*(4), 529-569. doi: 10.1207/s15327906mbr3804_5

Dart, B. C., & Clarke, J. A. (1991). Helping students become better learners: A case study in teacher education. *Higher Education, 22*(3), 317-335. doi: 10.1007/BF00132294

De Roche, K. K. (2009). *The functioning of global fit statistics in latent growth curve modeling.* Published Doctor of Philosophy dissertation, University of Northern Colorado.

Demirtas, H., & Schafer, J. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine, 22*, 2553-2575. doi: 10.1002/sim.1475

Desmedt, E., & Valcke, M. (2004). Mapping the learning styles "jungle": An overview of the literature based on citation analysis. *Educational Psychology, 24*(4), 2004. doi: 10.1080/0144341042000228843

Diseth, A. (2007). Approaches to learning, course experience and examination grade among undergraduate psychology students: testing of mediator effects and construct validity. *Studies in Higher Education, 32*(3), 373-388. doi: 10.1080/03075070701346949

DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment, 23*, 225-241. doi: 10.1177/073428290502300303

Donche, V., Coertjens, L., & Van Petegem, P. (2010). Learning pattern development throughout Higher Education: A longitudinal study. *Learning and Individual differences, 20*(3), 256-259. doi: 10.1016/j.lindif.2010.02.002

Donche, V., Coertjens, L., Vanthournout, G., & Van Petegem, P. (2012). Providing constructive feedback on learning patterns : An individual learner perspective. *Reflecting education, 8*, 114-132.

Donche, V., De Maeyer, S., Coertjens, L., van Daal, T., & Van Petegem, P. (2013). Differential use of learning strategies in first-year higher education: The impact of personality, academic motivation, and teaching strategies. *British Journal of Educational Psychology, 83*(2), 238–251. doi: 10.1111/bjep.12016

Donche, V., & Van Petegem, P. (2006). *Learning patterns in higher education: the influence of personal and contextual factors*. Paper presented at the British Educational Research Association Annual Conference, Warwick.

Donche, V., & Van Petegem, P. (2008). The validity and reliability of the Short Inventory of Learning Patterns. In E. Cools (Ed.), *Style and cultural differences: How can organisations, regions and countries take advantage of style differences.* (pp. 49-59). Gent, Belgium: Vlerick Leuven Gent Management School.

Donche, V., & Van Petegem, P. (2009). The development of learning patterns of student-teachers: A cross-sectional and longitudinal study. *Higher Education, 57*(4), 463-475. doi: 10.1007/s10734-008-9156-y

Donche, V., & Van Petegem, P. (2011). The relationship between entry characteristics, learning style and academic achievement of college freshmen. In M. E. Poulsen (Ed.), *Higher education: teaching, internationalization and student issues* (pp. 277-288). New York, NY: Nova Science Publishers.

Duncan, T., Duncan, S., & Strycker, L. (2006). *An introduction to latent variable growth curve modeling. Concepts, issues and applications*. London, United Kingdom: Erlbaum.

Endedijk, M. (2010). *Student teachers' self-regulated learning.* Utrecht University, Utrecht.

Endedijk, M. D., & Vermunt, J. D. (2013). Relations between student teachers' learning patterns and their concrete learning activities. *Studies in Educational Evaluation, 39*(1), 56-65. doi: 10.1016/j.stueduc.2012.10.001

Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement, 61*(5), 713-740. doi: 10.1177/0013164401615001

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: The Guilford Press.

Enders, C. K. (2011). Missing not at random models for latent growth curve analysis. *Psychological Methods, 16*(1), 1-16. doi: 10.1037/a0022640

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 430-457. doi: 10.1207/S15328007SEM0803_5

Entwistle, N., & McCune, V. (2004). The conceptual bases of study strategy inventories. *Educational Psychology Review, 16*, 325-345. doi: 10.1007/s10648-004-0003-0

Entwistle, N., McCune, V., & Hounsell, J. (2003). Investigating ways of enchancing university teaching-learning environments: Measuring students' approaches to studying and perceptions of teaching. In E. De Corte, L. Verschaffel, N. Entwistle & J. van Merriënboer (Eds.), *Powerful learning environments: Unraveling basic components and dimension* (pp. 89-109). Amsterdam: Pergamon.

Entwistle, N., McCune, V., & Scheja, M. (2006). Student learning in context: Understanding the phenomenon and the person. In L. Verschaffel, F. Dochy, M. Boekaerts & S. Vosniadou (Eds.), *Instructional Psychology: Past, Present and Future Trends* (pp. 131-148). Amsterdam: Elsevier.

Entwistle, N., Meyer, J., & Tait, H. (1991). Student failure: disintegrated perceptions of study strategies and perceptions of the learning environment. *Higher Education, 21*, 249-261. doi: 10.1007/BF00137077

Entwistle, N., & Ramsden, P. (1983). *Understanding Student Learning*. London, United Kingdom: Croom Helm.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis; Verbal reports as data* (revised ed.). Cambridge, MA: Bradford books/MIT Press.

European Commission/EACEA/Eurydice. (2014). Modernisation of Higher Education in Europe. Access, Retention and Employability 2014. Eurydice Report. Luxembourg: Publications Office of the European Union.

Fidalgo, A. M., & Scalon, J. D. (2010). Using generalized Mantel-Haenszel statistics to assess DIF among multiple groups. *Journal of Psychoeducational Assessment, 28*(1), 60-69. doi: 10.1177/0734282909337302

Flavell, J. (1987). Speculations about the nature and development of metacognition. In F. Weinert & R. Kluwe (Eds.), *Metacognition, motivation and understanding* (pp. 21-29). Hillsdale, New Jersey: Erlbaum.

Fleming, C. B., Catalano, R. F., Mazza, J. J., Brown, E. C., Haggarty, K. P., & Harachi, R. W. (2008). After-school activities, misbehavior in school and delinquency from the end of elementary school through the beginning of high school: A test of social development model hypotheses. *Journal of Early Adolescence, 28*(2), 277-303. doi: 10.1177/0272431607313589

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466-491. doi: 10.1037/1082-989X.9.4.466

Foster, E. M., Fang, G. Y., & Conduct Problems Prevention Research Group. (2004). Alternative methods for handling attrition: An illustration using data from the fast track evaluation. *Evaluation Review, 28*(5), 434-464. doi: 10.1177/0193841X04264662

Fox, A., Stevenson, L., Connelly, P., Duff, A., & Dunlop, A. (2010). Peer-mentoring undergraduate accounting students: The influence on approaches to learning and academic performance. *Active Learning in Higher Education, 11*(2), 145–156. doi: 10.1177/1469787410365650

Gent, I., Johnston, B., & Prosser, P. (1999). Thinking on your feet in undergraduate computer science: A constructivist approach to developing and assessing critical thinking. *Teaching in Higher Education, 4*(4), 511-522. doi: 10.1080/1356251990040407

Germeijs, V., & Verschueren, K. (2007). High school students' career decision-making process: Consequences for choice implementation in higher education. *Journal of Vocational Behavior, 70*, 223-241. doi: 10.1016/j.jvb.2006.10.004

Gibbons, R. D., Hedeker, D., Elkin, I., Waternaux, C., Kraemer, H. C., Greenhouse, J. B., et al. (1993). Some conceptual and statistical issues in analysis of longitudinal psychiatric data: Application to the nimh treatment of depression collaborative research program dataset. *Archives of General Psychiatry, 50*(9), 739-750. doi: 10.1001/archpsyc.1993.01820210073009

Gijbels, D., Coertjens, L., Vanthournout, G., Struyf, E., & Van Petegem, P. (2009). Changing students' approaches to learning: A two year study within a university teacher training course. *Educational Studies, 35*, 503-513. doi: 10.1080/03055690902879184

Gijbels, D., Raemdonck, I., & Vervecken, D. (2010). Influencing work-related learning: The role of job characteristics and self-directed learning orientation in part-time vocational education. *Vocations and Learning, 3*, 329-255. doi: 10.1007/s12186-010-9041-6

Gordon, C., & Debus, R. (2002). Developing deep learning approaches and personal teaching efficacy within a preservice teacher education context. *British Journal of Educational Psychology, 72*, 483-511. doi: 10.1348/00070990260377488

Gottman, J. M. (1995). *Analysis of change*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gow, L., & Kember, D. (1990). Does higher education promote independent learning? *Higher Education, 19*, 307-322. doi: 10.1007/BF00133895

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549-576. doi: 10.1146/annurev.psych.58.110405.085530

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8*, 206-213. doi: 10.1007/s11121-007-0070-9

Graham, J. W., & Schafer, J. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 1-29). Thousand Oaks, CA: Sage.

Green, S. B., & Salkind, N. J. (2003). *Using SPSS for Windows and Macintosh. Analyzing and Understanding Data.* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Grimm, K. J., & Ram, N. (2009). Nonlinear growth models in Mplus and SAS. *Structural Equation Modeling, 16*(4), 676-701. doi: 10.1080/10705510903206055

Guttmannova, K., Szanyi, J. M., & Cali, P. W. (2008). Internalizing and externalizing behavior problem scores: Cross-ethnic and longitudinal measurement invariance of the Behavior Problem Index. *Educational and Psychological Measurement, 68*(4), 676-694. doi: 10.1177/0013164407310127

Heck, R. H., Scott, L. T., & Tabata, L. N. (2010). *Multilevel and longitudinal modeling with IBM SPSS*. New York, NY: Routledge.

Hedeker, D., & Gibbons, R. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2*(1), 64-78. doi: 10.1037/1082-989X.2.1.64

Hoeksema, L., van de Vliert, E., & Williams, R. (1997). The interplay between learning strategy and organizational structure in predicting career success. *The International Journal of Human Resource Management, 8*(3), 307-327. doi: 10.1080/095851997341667

Hox, J. J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 147-154). New York, NY: Springer Verlag.

Hox, J. J. (2000). Multilevel analysis of grouped and longitudinal data. In T. D. Little, K. U. Schabel & J. Baumert (Eds.), *Modeling longitudinal and multi-group data: Practical issues, applied approaches, and specific examples*. Hillsdale, NJ: Erlbaum.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indixes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. doi: 10.1080/10705519909540118

Hultberg, J., Plos, K., Hendry, G. D., & Kjellgren, K. I. (2008). Scaffolding students' transition to higher education: Parallel introductory courses for students and teachers. *Journal of Further and Higher Education, 32*(1), 47-57. doi: 10.1080/03098770701781440

Jackling, B. (2005). Analysis of the learning context, perceptions of the learning environment and approaches to learning accounting: A longitudinal study. *Accounting and Finance, 45*, 597-612. doi: 10.1111/j.1467-629X.2005.00142.x

Jeličić, H., Phelps, E., & Lerner, R. M. (2010). Why missing data matter in the longitudinal study of adolescent development: Using the 4-H study to understand the uses of different missing data methods. *Journal of Youth Adolescence, 39*, 816-835. doi: 10.1007/s10964-010-9542-5

Johnson, D. R., & Young, R. (2011). Toward best practices in analyzing datasets with missing data: Comparisons and recommendations. *Journal of Marriage and Family, 73*, 926-945. doi: 10.1111/j.1741-3737.2011.00861.x

Jung, T., & Wickrama, K. (2008). An introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass, 2*(1), 302-317. doi: 10.1111/j.1751-9004.2007.00054.x

Kember, D., & Leung, D. Y. P. (1998). The dimensionality of approaches to learning: An investigation with confirmatory factor analysis on the structure of the SPQ and LPQ. *British Journal of Educational Psychology, 68*, 395-407. doi: 10.1111/j.2044-8279.1998.tb01300.x

Kim, S.-Y., & Kim, J.-S. (2012). Investigating stage-sequential growth mixture models with multiphase longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(2), 293-319. doi: 10.1080/10705511.2012.659632

Kolb, D. (1984). *Experiential Learning. Experience as a source of learning and development.* Englewood Cliffs, New Jersey: Prentice Hall Inc.

Kreber, C. (2003). The Relationship between Students' Course Perception and their Approaches to Studying in Undergraduate Science Courses: A Canadian experience. *Higher Education Research & Development, 22*(1), 57-75. doi: 10.1080/0729436032000058623

Kyndt, E., Dochy, F., Cascallar, E., & Struyven, K. (2011). The direct and indirect effect of motivation for learning on students' approaches to learning through the perceptions of workload and task complexity. *Higher Education Research and Development, 30*(2), 135-150. doi: 10.1080/07294360.2010.501329.

Li, F., Duncan, T. E., Duncan, S. C., & Hops, H. (2001). Piecewise growth mixture modeling of adolescent alcohol use data. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(2), 175-204. doi: 10.1207/S15328007SEM0802_2

Li, F., Harmer, P., & Acock, A. (1996). The task and ego orientation in sport questionnaire: construct equivalence and mean differences across gender. *Research Quarterly for Exercise and Sport, 67*(2), 228-238. doi: 10.1080/02701367.1996.10607949

Lindblom-Ylänne, S. (2003). Broadening understanding of the phenomenon of dissonance. *Studies in Higher Education, 28*, 63-77. doi: 10.1080/03075070309306

Lindblom-Ylänne, S., & Lonka, K. (1999). Individual ways of interacting with the learning environment - Are they related to study success? *Learning and Instruction, 9*(1), 1-18. doi: 10.1016/S0959-4752(98)00025-5

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (Second edition ed.). Hoboken, NJ: Wiley Interscience.

Lonka, K., Olkinuora, E., & Mäkinen, J. (2004). Aspects and Prospects of Measuring Studying and Learning in Higher Education. *Educational Psychology Review, 16*(4), 301-323. doi: 10.1007/s10648-004-0002-1

Mäkinen, J., Olkinuora, E., & Lonka, K. (2004). Students at risk: Students' general study orientations and abandoning/prolonging the course of studies. *Higher Education, 48*(2), 173-188. doi: 10.1023/B:HIGH.0000034312.79289.ab

Marambe, K. (2007). *Patterns of student learning in Medical Education - A Sri Lankan study in a traditional curriculum*. Unpublished doctoral dissertation. University of Maastricht. Maastricht.

Marsh, H. W., & Grayson, D. (1994). Longitudinal stability of latent means and individual differences: A unified approach. *Structural Equation Modeling, 1*(2), 317-359. doi: 10.1080/10705519409539984

Marsh, H. W., & Hau, K.-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology, 32*(1), 151-170. doi: 10.1016/j.cedpsych.2006.10.008

Marsh, H. W., Hau, K.-T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's Brief Self-Report Measure of Educational Psychology's Most Useful Affective Constructs: Cross-Cultural, Psychometric Comparisons Across 25 Countries. *International Journal of Testing, 6*(4), 311-360. doi: 10.1207/s15327574ijt0604_1

Marton, F., & Säljö, R. (1976). On qualitative differences in learning - I: Outcome and process. *British Journal of Educational Psychology, 46*, 4-11. doi: 10.1111/j.2044-8279.1976.tb02980.x

Marton, F., & Säljö, R. (1997). Approaches to learning. In F. Marton, D. Hounsell & N. J. Entwistle (Eds.), *The Experience of Learning. Implications for Teaching and Studying in Higher Education*. Edinburgh: Scottish Academic Press.

Mayer, R. E. (2011). Does styles research have useful implications for educational practice? *Learning and Individual Differences, 21*, 319-320. doi: 10.1016/j.lindif.2010.11.016

McArdle, J. J., & Nesselroade, J. R. (2013). An overview of latent curve and latent change score analyses. In J. A. Schinka, W. F. Velicer & I. B. Weiner (Eds.), *Handbook of psychology: Volume 2 Research methods in psychology*. Hoboken, NJ: John Wiley & Sons.

McManus, I. C., Keeling, A., & Paice, E. (2004). Stress, burnout and doctors' attitudes to work are determined by personality and learning style: A twelve year longitudinal study of UK medical graduates. *BMC Medicine, 2*(1), 1-12. doi: 10.1186/1741-7015-2-29

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568-592. doi: 10.1037/0021-9010.93.3.568

Messick, S. (1996). Cognitive styles and learning. In E. De Corte & F. Weinert (Eds.), *International encyclopaedia of developmental and instructional psychology* (pp. 638 – 641). Oxford: Elsevier Science Ltd,.

Metha, P. D., Neale, M. C., & Flay, B. R. (2004). Squeezing interval change from ordinal panel data: Latent growth curves with ordinal outcomes. *Psychological Methods, 9*(3), 301-333. doi: 10.1037/1082-989X.9.3.301

Meyer, J. H. F. (1998). A Medley of Individual Differences. In B. Dart & G. Boulton-Lewis (Eds.), *Teaching and Learning in Higher Education*. Melbourne: ACER-Press.

Molenberghs, G., & Fitzmaurice, G. (2009). Incomplete data: Introduction and overview. In G. Fitzmaurice, M. Davidian, G. Verbeke & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 395-408). Boca Raton, FL: CRC Press.

Moskvina, V., & Kozhevnikov, M. (2011). Determining cognitive styles: Historical perspective and directions for further research. In S. Rayner & E. Cools (Eds.), *Style differences in cognition, learning, and management: Theory, research, and practice*. New York: Routledge.

Muthén, B., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods, 16*(1), 17-33. doi: 10.1037/a0022634

Muthén, B. O. (2007). Non-positive definite psi matrix. Retrieved 20 May 2011, from http://www.statmodel.com/discussion/messages/14/815.html?1279923698

Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2*, 371-402. doi: 10.1037/1082-989x.2.4.371

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Retrieved from http://pages.gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf

Muthén, L. K., & Muthén, B. O. (2009a). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(4), 599-620. doi: 10.1207/S15328007SEM0904_8

Muthén, L. K., & Muthén, B. O. (2009b). *Mplus user's guide* (5th ed.). Los Angeles, CA: Author.

Muthén, L. K., & Muthén, B. O. (2010). *Growth modeling with latent variable using Mplus: Advanced growth models, survival analysis and missing data*. Mplus Short Courses.

Nicholson, N., & West, M. (1995). Transitions, work histories, and careers. In M. B. Arthur, D. T. Hall & B. S. Lawrence (Eds.), *Handbook of career theory* (pp. 181-201). Cambridge: Cambridge University Press.

Nieminen, J., Lindblom-Ylänne, S., & Lonka, K. (2004). The development of study orientations and study success in students of pharmacy. *Instructional Science, 32*, 387-417.

Nienemin, J., Lindblom-Ylänne, S., & Lonka, K. (2004). The development of study orientations and study success in students of pharmacy. *Instructional Science, 32*, 387-417. doi: 10.1023/B:TRUC.0000044642.35553.e5

Nussbeck, F. W., Eid, M., & Lischetzke, T. (2006). Analysing multitrait-multimethod data with structural equation models for ordinal variables applying the WLSMV estimator: What sample size is needed for valid results? *British journal of Mathematical and Statistical Psychology, 59*, 195-213. doi: 10.1348/000711005X67490

OECD. (2013). Education at a glance: OECD indicators.

Palant, J. (2007). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows* (3rd ed.). New York, NY: Open University Press.

Parpala, A., Lindblom-Ylänne, S., Komulainen, E., Litmanen, T., & Hirsto, L. (2010). Students' approaches to learning and their experiences of the teaching-learning environment in different disciplines. *British Journal of Educational Psychology, 80*(2), 269-282. doi: 10.1348/000709909X476946

Petscher, Y., & Huijun, L. (2008). Measurement invariance of the Chinese Gifted Rating Scales: Teacher and parent forms. *Journal of Psychoeducational Assessment, 26*(3), 274-286. doi: 10.1177/0734282907303873

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*(4), 525-556. doi: 10.3102/00346543074004525

Phan, H. P. (2008). Predicting change in epistemological beliefs, reflective thinking and learning styles: A longitudinal study. *British Journal of Educational Psychology, 78*, 75-93. doi: 10.1348/000709907X204354

Phan, H. P. (2011). Deep processing strategies and critical thinking: Developmental trajectories using latent growth analyses. *The Journal of Educational Research, 104*(4), 283-294. doi: 10.1080/00220671003739382

Provost, S. C., & Bond, N. W. (1997). Approaches to studying and academic performance in a traditional psychology course. *Higher Education Research and Development, 16*(3), 309-320. doi: 10.1080/0729436970160305

Quené, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech communication, 43*(1-2), 103-121. doi: 10.1016/j.specom.2004.02.004

Raudenbush, S. W. (2001a). Comparing personal trajcetories and drawing causal inferences from longitudinal data. *Annual Review of Psychology, 52*, 501-525. doi: 10.1146/annurev.psych.52.1.501

Raudenbush, S. W. (2001b). Toward a coherent framework for comparing trajectories of individual change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 35-64). Washington, DC: American Psychological Association.

Reid, W. A., Duvall, E., & Evans, P. (2005). Can we influence medical students' approaches to learning? *Medical Teacher, 27*(5), 401-407. doi: 10.1080/01421590500136410

Richardson, J. (1997). Meaning orientation and reproducing orientation: A typology of approaches to studying in higher education? *Educational Psychology, 17*, 301-311.

Richardson, J. T. E. (2004). Methodological Issues in Questionnaire-Based Research on Student Learning in Higher Education. *Educational Psychology Review, 16*(4), 347-358. doi: 10.1007/s10648-004-0004-z

Richardson, J. T. E. (2006). Perceptions of academic quality and approaches to studying among technology students in distance education. *European Journal of Engineering Education, 31*(4), 421-433. doi: 10.1080/03043790600676307

Richardson, J. T. E. (2007). Variations in student learning and perceptions of academic quality. In N. Entwistle & P. Tomlinson (Eds.), *Student Learning and University Teaching* (pp. 61-71). Leicester: The British Psychological Society.

Richardson, J. T. E. (2011). Approaches to studying, conceptions of learning and learning styles in higher education. *Learning and Individual Differences, 21*, 288-293. doi: 10.1016/j.lindif.2010.11.015

Richardson, J. T. E. (2013). Research issues in evaluating learning pattern development in higher education. *Studies in Educational Evaluation, 39*, 66-70. doi: 10.1016/j.stueduc.2012.11.003

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353-387. doi: 10.1037/a0026838

Rocconi, L., & Ethington, C. (2009). Assessing longitudinal change: Adjustment for regression to the mean effects. *Research in Higher Education, 50*, 368-376. doi: 10.1007/s11162-009-9119-x

Rozendaal, J. S., Minnaert, A., & Boekaerts, M. (2003). Motivation and self-regulated learning in secondary vocational education: Information-processing type and gender differences. *Learning and Individual Differences, 13*, 273-289. doi: 10.1016/S1041-6080(03)00016-5

Sadler-Smith, E. (1996). Approaches to studying: Age, gender and academic performance. *Educational Studies, 22*, 367-379. doi: 10.1080/0305569960220306

Sadler-Smith, E., & Tsang, F. (1998). A comparative study of approaches to studying in Hong Kong and the United Kingdom. *British Journal of Educational Psychology, 68*, 81-93. doi: 10.1111/j.2044-8279.1998.tb01276.x

Sander, P., Stevenson, K., King, M., & Coates, D. (2000). University Students' Expectations of Teaching. *Studies in Higher Education, 25*(3), 309-323. doi: 10.1080/03075070050193433

Schatteman, A., Carette, E., Couder, J., & Eisendrath, H. (1997). Understanding the effects of a process-oriented instruction in first year of university by investigating learning style characteristics. *Educational Psychology, 17*(1), 111-125. doi: 10.1080/0144341970170108

Schmeck, R. (1988). An introduction to Strategies and Styles of Learning. In R. Schmeck (Ed.), *Learning Strategies and Learning Styles*. New York: Plenum Press.

Schmitt, M. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350-353. doi: 10.1037/1040-3590.8.4.350

Schofer, E., & Meyer, J. W. (2005). The worldwide expansion of higher education in the twentieth century. *American Sociological Review, 70*(6), 898-920. doi: 10.1177/000312240507000602

Schünemann, N., Spörer, N., & Brunstein, J. C. (2013). Integrating self-regulation in whole-class reciprocal teaching: A moderator–mediator analysis of incremental effects on fifth graders' reading comprehension. *Contemporary Educational Psychology, 38*(4), 289-305. doi: 10.1016/j.cedpsych.2013.06.002

Segers, M., Nijhuis, J., & Gijselaers, W. (2006). Redesigning a learning and Assessment Environment: the Influence on Students' Perceptions of Assessment Demands and their Learning Strategies. *Studies in Educational Evaluation, 32*(223-242). doi: l 0.1016/j.stueduc.2006.08.004

Severiens, S., & Ten Dam, G. (1998). Gender and learning: Comparing two theories. *Higher Education, 35*(3), 329-350. doi: 10.1023/a:1003033003124

Severiens, S., Ten Dam, G., & Van Hout-Wolters, B. (2001). Stability of processing and regulation strategies: Two longitudinal studies on student learning. *Higher Education, 42*, 437-453. doi: 10.1023/A:1012227619770

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Shin, J. H., Braines, R. B., & Johnston, M. E. (1993). Effect of problem-based, self-directed undergraduate education on life-long learning. *Canadian Medical Association Journal, 148*, 969-976.

Singer, J., & Willet, J. (2003). *Applied longitudinal data analysis. Modelling change and event occurrence.* Oxford, United Kingdom: University Press.

Smith, L., Saini, B., Krass, I., Chen, T., Bosnic-Anticevich, S., & Sainsbury, E. (2007). Pharmacy students' approaches to learning in an Australian University. *American Journal of Pharmaceutical Education, 71*(6), Article 120. doi: 10.5688/aj7106120

Steinmetz, H., Schmidt, P., Tina-Booth, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality & Quantity, 43*(4), 599-616. doi: 10.1007/s11135-007-9143-x

Stevenson, K., Sander, P., & Naylor, P. (1996). Student perceptions of the tutor's role in distance learning. *Open Learning: The Journal of Open, Distance and e-Learning, 11*(1), 22-30. doi: 10.1080/0268051960110103

Stoel, R. D., Roeleveld, J., Peetsma, T., van den Wittenboer, G., & Hox, J. (2006). Issues in longitudinal research on motivation. *Learning and Individual differences, 16*(3), 159-174. doi: 10.1016/j.lindif.2005.07.006

Stoel, R. D., & van Den Wittenboer, G. (2003). Analyzing longitudinal data using multilevel regression and latent growth curve analysis. *Metodología de las ciencias del comportamiento, 5*(1), 21-42.

Struyven, K., Dochy, F., Janssens, S., & Gielen, S. (2006). On the dynamics of students' approaches to learning: The effects of the teaching/learning environment. *Learning and Instruction, 16*, 279-294. doi: 10.1016/j.learninstruc.2006.07.001

Tait, H., Entwistle, N. J., & McCune, V. (1998). ASSIST: A reconceptualisation of the Approaches to Studying Inventory In C. Rust (Ed.), *Improving student learning: Improving students as learners* (pp. 262-271). Oxford, United Kingdom: Oxford Centre for Staff and Learning Development.

The European Commission. (2010). *The European Qualifications Framework for Lifelong Learning*. Retrieved from http://ec.europa.eu/eqf/documentation_en.htm.

Tomarken, A. J., & Waller, N. G. (2005). Structural Equation Modeling: Strengths, Limitations, and Misconceptions. *Annual Review of Clinical Psychology, 1*(1), 31-65. doi: 10.1146/annurev.clinpsy.1.102803.144239

Tooth, D., Tonge, K., & McManus, I. C. (1989). Anxiety and study methods in preclinical students: causal relation to examination performance. *Medical Education, 23*(5), 416-421. doi: 10.1111/j.1365-2923.1989.tb00896.x

Torenbeek, M., Jansen, E., & Hofman, A. (2009). How first year students perceive the fit between secondary and university education: The effect of teaching approaches. *Effective Education, 1*(2), 135-150. doi: 10.1080/19415530903522543

Torenbeek, M., Jansen, E. P. W. A., & Hofman, W. H. A. (2010). The effect of the fit between secondary and university education of first-year student achievement. *Studies in Higher Education, 35*(6), 659-675. doi: 10.1080/03075070903222625

Trigwell, K., Prosser, M., & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education, 37*, 57-70. doi: 10.1023/A:1003548313194

Vaillancourt, T., Brengden, M., Boivin, M., & Tremblay, R. E. (2003). A longitudinal confirmatory factor analysis of indirect and physical aggression: Evidence of two factors over time. *Child Development, 74*(6), 1628-1638. doi: 10.1046/j.1467-8624.2003.00628.x

van Bragt, C. A. C., Bakx, A. W. E. A., Van der Sanden, J. M. M., & Croon, M. E. (2007). Students' approaches to learning when entering higher education:

Differences between students with senior general secondary and senior secondary educational backgrounds. *Learning and Individual Differences, 17*, 83-96. doi: 10.1016/j.lindif.2007.02.003

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486-492. doi: 10.1080/17405629.2012.686740

van der Veen, I., & Peetsma, T. (2009). The development of self-regulated learning behavior of first-year students in the lowest level of secondary school in the Netherlands. *Learning and Individual Differences, 19*, 34-46. doi: 10.1016/j.lindif.2008.03.001

Van der Veken, J., Valcke, M., De Maeseneer, J., & Derese, A. (2009). Impact of the transition from a conventional to an integrated contextual medical curriculum on students' learning patterns: A longitudinal study. *Medical Teacher, 31*, 433-441. doi: 10.1080/01421590802141159

Van Rossum, E., & Schenk, S. (1984). The relationship between learning conception, study strategy and learning outcome. *British Journal of Educational Psychology, 54*, 73-83. doi: 10.1111/j.2044-8279.1984.tb00846.x

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-69. doi: 10.1177/109442810031002

Vanthournout, G. (2011). *Patterns in student learning: Exploring a person-oriented and longitudinal research-perspective*. Antwerp: Garant.

Vanthournout, G., Donche, V., Gijbels, D., & Van Petegem, P. (2011). Further understanding learning in higher education: A systematic review on longitudinal research using Vermunt's learning pattern model In S. Rayner & E. Cools (Eds.), *Style differences in cognition, learning and management: Theory, research and practice* (pp. 78-96). London, United Kingdom: Routledge.

Vanthournout, G., Donche, V., Gijbels, D., & Van Petegem, P. (2014). (Dis)similarities in research on learning approaches and learning patterns. In D. Gijbels, V. Donche, J. T. E. Richardson & J. D. Vermunt (Eds.), *Learning patterns in higher education. Dimensions and research perspectives*. London: Routledge.

Vanthournout, G., Gijbels, D., Coertjens, L., Donche, V., & Van Petegem, P. (2012). Students' persistence and academic success in a first-year professional bachelor program: The influence of students' learning strategies and academic motivation. *Education Research International*, Article ID 152747. doi: 10.1155/2012/152747

Vermetten, Y., Lodewijks, H. G. L. C., & Vermunt, J. D. (1999a). Consistency and variability of learning strategies in different university courses. *Higher Education, 37*(1), 1-21. doi: 10.1023/A:1003573727713

Vermetten, Y., Vermunt, J. D., & Lodewijks, H. (1999b). A longitudinal perspective on learning strategies in higher education: Different viewpoints towards development. *British Journal of Educational Psychology, 69*(2), 221-242. doi: 10.1348/000709999157699

Vermetten, Y., Vermunt, J. D., & Lodewijks, H. (2002). Powerfull learning environments? How do students differ in their response to instructional measures. *Learning and Instruction, 12*, 263-284. doi: 10.1016/S0959-4752(01)00013-5

Vermunt, J. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher Education, 31*(1), 25-50. doi: 10.1007/BF00129106

Vermunt, J. (1998). The regulation of constructive learning processes. *British Journal of Educational Psychology, 68*, 149-171. doi: 10.1111/j.2044-8279.1998.tb01281.x

Vermunt, J. (2005). Relations between student learning patterns and personal and contextual factors and academic performance. *Higher Education, 49*, 205-234. doi: 10.1007/s10734-004-6664-2

Vermunt, J., & Minnaert, A. (2003). Dissonance in student learning patterns: When to revise theory? *Studies in Educational Evaluation, 28*, 49-61. doi: 10.1080/03075070309301

Vermunt, J., & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction, 9*, 257-280. doi: 10.1016/S0959-4752(98)00028-0

Vermunt, J., & Vermetten, Y. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning and learning orientations. *Educational Psychology Review, 16*(4), 359 - 384. doi: 10.1007/s10648-004-0005-y

Vermunt, J. D., Richardson, J. T. E., Donche, V., & Gijbels, D. (2014). Students' learning patterns in higher education. Dimensions, measurement and change. In D. Gijbels, V. Donche, J. T. E. Richardson & J. D. Vermunt (Eds.), *Learning patterns in Higher Education. Dimensions and research perspectives*. London: Routledge.

Voelkle, M. C., Wittmann, W. W., & Ackerman, P. L. (2006). Abilities and skill acquisition: A latent growth curve approach. *Learning and Individual Differences, 16*, 303-319. doi: 10.1016/j.lindif.2006.01.001

Volet, S. E., Renshaw, P. D., & Tietzel, K. (1994). A short-term longitudinal investigation of cross-cultural differences in study approaches using Biggs' SPQ questionnaire. *British Journal of Educational Psychology, 64*, 301-318. doi: 10.1111/j.2044-8279.1994.tb01104.x

Wang, J., & Wang, X. (2012). *Structural equation modeling. Applications using Mplus*. West Sussex: Wiley.

Watkins, D., & Hattie, J. (1985). A longitudinal study of the approaches to learning of Australian tertiary students. *Human Learning: Journal of Practical Research & Applications, 4*, 127-141.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604. doi: 10.1037/0003-066X.54.8.594

Wilson, K., & Fowler, J. (2005). Assessing the impact of learning environments on students' approaches to learning. *Assessment and Evaluation in Higher Education, 30*(1), 37-56.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T. D. Little, K. U. Schnabel & J. Baumert (Eds.), *Modeling longitudinal and multilevel data. Practical issues, applied approaches, and specific examples*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wu, A., Liu, Y., Gadermann, A. M., & Zumbo, B. D. (2010). Multiple-indicator multilevel growth model: A solution to multiple methodological challenges in longitudinal studies. *Social Indicators Research, 97*(2), 123-142. doi: 10.1007/s11205-009-9496-8

Xu, S., & Blozis, S. A. (2011). Sensitivity analysis of mixed models for incomplete longitudinal data. *Journal of Educational and Behavioral Statistics, 36*(2), 237-256. doi: 10.3102/1076998610375836

Yoo, L. E. (2009). The effect of auxiliary variables and multiple imputation on parameter estimation in confirmatory factor analysis. *Educational and Psychological Measurement, 69*(6), 929-947. doi: 10.1177/0013164409332225

Zeegers, P. (2001). Approaches to learning in science: A longitudinal study. *British Journal of Educational Psychology, 71*(1), 115-132. doi: 10.1348/000709901158424

Zhang, L., & Sternberg, R. J. (2005). A Threefold Model of Intellectual Styles. *Educational Psychology Review, 17*, 1-53. doi: 10.1007/s10648-005-1635-4

Appendix: Glossary

Term	Defined as	For more information, see section
Approach	the combination of a strategy (how a student learns) and a motivational component (why a student learns)	1.1.1.
Average growth	Whether and how students, on average, change in a construct	1.2.
Change	Significant increase or decrease in a construct; synonym: growth	1.2.
Differential growth	Whether and how students vary in their growth over time	1.2.
ILS-SV	Inventory of Learning Styles – Short Version; self-report questionnaire with 30 items mapping four processing and three regulation strategies	
Learning pattern	The combination of a processing, regulation, motivational component (named orientation to learning) and learning conception	1.1.1
Learning pattern model	A model in the SAL tradition which describes differences in student learning by means of learning patterns	Table 1.1 in 1.1.2.
Learning strategy	The way a student habitually goes about learning. The SAL model views learning strategies as processing strategies while the learning patttern model views learning strategies as processing and regulation strategies	1.1.2.
Processing strategy	The cognitive activities a student usually applies whilst studying	1.1.2.
Regulation strategy	The meta-cognitive activities students usually rely on to direct their learning process	1.1.2.
Growth	Significant change in a construct over time, regardless of whether this change is judged as positive; synonym: change	1.2.
Growth trend	The average and differential growth over time	1.2.
SAL tradition	Tradition of thinking about learning founded by Marton & Säljö which states that students differ in how they go about learning. These differences affect how well students process learning content.	1.1.
SAL model	A model in the SAL tradition in which learning strategies are viewed as processing strategies	1.1.1

Dutch summary / Nederlandstalige samenvatting

Verandering in leerstrategieën tijdens en na de overstap van het secundair onderwijs naar het hoger onderwijs: Hoe beïnvloeden statistische keuzes de parameterschattingen?

Studenten voorbereiden op levenslang leren is een belangrijke uitdaging voor het onderwijs. Zelfregulatie en diepe verwerkingsstrategieën zijn belangrijke vaardigheden om dit levenslang leren te bevorderen. Om deze reden evolueert de manier waarop studenten gewoonlijk het leren aanpakken, ofwel hun 'leerstrategieën', idealiter in de richting van diepere verwerking en meer zelf-gereguleerd leren. Dit proefschrift beschrijft hoe leerstrategieën veranderen tijdens en na de overstap van het secundair onderwijs naar het hoger onderwijs en gaat na hoe statistische keuzes de parameterschattingen van die verandering beïnvloeden.

Inleiding

Enkele studies focusten reeds op hoe de leerstrategieën van studenten veranderen doorheen de tijd. Er zijn echter twee grote tekortkomingen aan deze studies. Ten eerste werden ze allen uitgevoerd bij studenten in het hoger onderwijs. Echter, de evolutie van leerstrategieën bekijken binnen één stabiele onderwijscontext vertekent mogelijk het beeld van de veranderlijkheid in leerstrategieën. Onderzoek naar de evolutie in leerstrategieën tijdens de overstap van secundair naar hoger onderwijs kan meer inzicht geven in hoe leerstrategieën evolueren doorheen de tijd: zijn er periodes waarin leerstrategieën redelijk stabiel blijven of waarin de verandering van leerstrategieën in een stroomversnelling zit?

Verder wordt er in eerder onderzoek naar de verandering in leerstrategieën vaak gebruik gemaakt van zelfrapportage vragenlijsten. Leerstrategieën worden hierin bevraagd aan de hand van een aantal items met een Likert antwoordschaal. Wanneer de verandering in leerstrategieën doorheen de tijd wordt onderzocht, moet bij de analyse van de longitudinale data een aantal statistische keuzes worden gemaakt. Hierin schuilt een tweede tekortkoming van de huidige studies. Zo wordt er in eerder onderzoek gewerkt met somscores in plaats van de items die een schaal uitmaken expliciet mee te nemen in de analyse. Bij gevolg wordt de error of ruis genegeerd die een item onvermijdelijk meet naast het eigenlijke concept.

Daarnaast vereist het in kaart brengen van verandering longitudinale meetinvariantie. Deze assumptie stelt dat het meetinstrument op verschillende dataverzamelingsmomenten op een vergelijkbare manier meet. Deze assumptie wordt echter niet getoetst in reeds gevoerd onderzoek, waardoor verandering in de meetlat mogelijk wordt verward met verandering in de leerstrategie.

Een laatste element betreft de studenten die uitvallen of niet antwoorden. Zij worden meestal niet opgenomen in de analyse, wat de bekomen resultaten kan vertekenen. Voor elk van deze drie statistische keuzes zijn er meer gepaste alternatieve keuzes beschreven in methodologische literatuur en beschikbaar in statistische software. Wat echter ontbreekt in de literatuur over het leren van studenten, is onderzoek dat deze alternatieve keuzes afzet tegen de gangbare aanpak en illustreert welke invloed ze hebben op de parameterschattingen.

Focus van het onderzoek

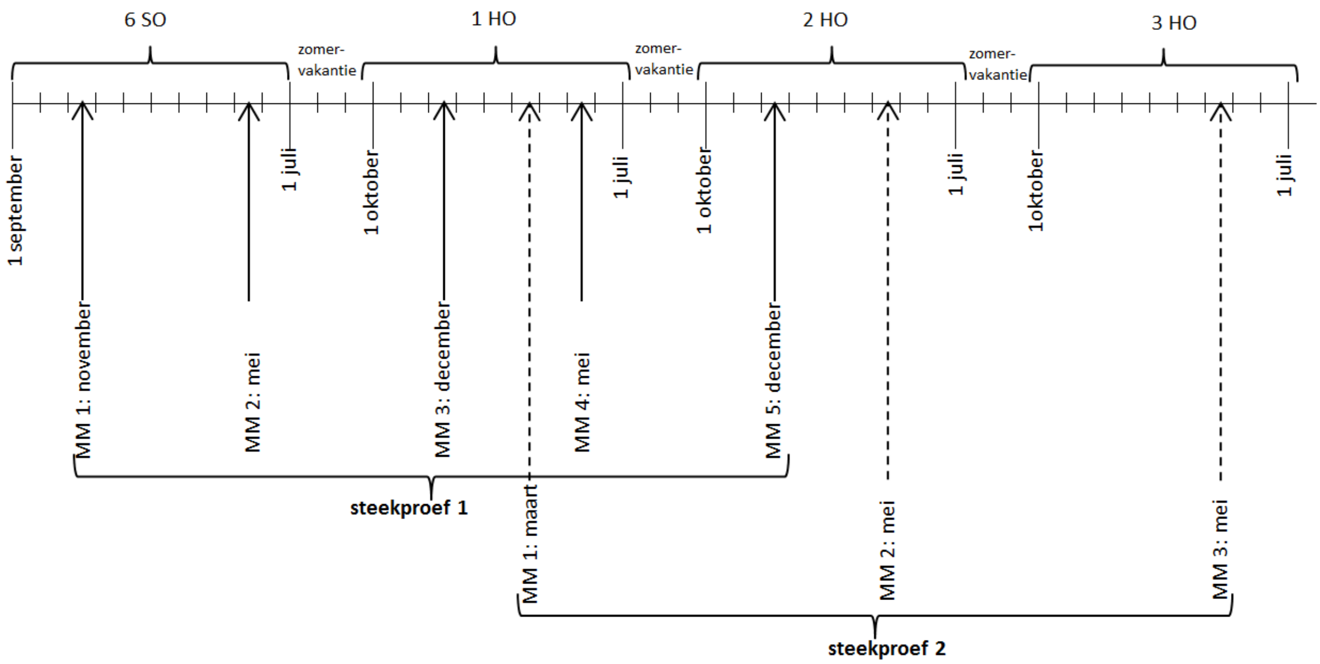
Het doel van dit proefschrift is tweeledig. Enerzijds brengt het in kaart hoe de leerstrategieën van studenten veranderen tijdens de overstap naar het hoger onderwijs en tijdens het hoger onderwijs. Dit komt aan bod in hoofdstukken 4 en 5. Anderzijds wordt inzicht gegeven in hoe de drie statistische keuzes de parameterschattingen van de verandering kunnen beïnvloeden. Hoofdstuk 2 bekijkt het effect van het meenemen van de items in de analyse als alternatief voor het gebruik van somscores. In hoofdstuk 3 wordt toegelicht hoe de assumptie van longitudinale meetinvariantie kan getoetst worden. Hoofdstuk 6, tenslotte, beschrijft de impact van de keuze van techniek om met ontbrekende data om te gaan op de schattingen over verandering in leerstrategieën.

Voor dit proefschrift werd gebruik gemaakt van de Leer- en Motivatie (LEMO) vragenlijst. Onderstaande tabel verduidelijkt de betekenis van de zeven leerstrategieën uit deze vragenlijst. De LEMO brengt vier verwerkingsstrategieën in kaart, namelijk relateren en structureren en kritisch verwerken, die beide diepe verwerkingsstrategie vatten, en memoriseren en analyseren, die een stapsgewijze manier van verwerken van de leerstof meten. Daarnaast brengen drie schalen in kaart hoe het leren van studenten wordt aangestuurd (regulatiestrategieën): zelfregulatie, externe regulatie en stuurloos leergedrag.

Tabel 1: Leerstrategieën uit de LEMO en hun betekenis

Schaal	Betekenis
Verwerkingsstrategieën	
Diepe verwerking	
Relateren en structureren	Mate waarin studenten naar verbanden zoeken in de leerinhoud
Kritisch verwerken	Mate waarin studenten kritisch staan tegenover de leerinhoud
Stapsgewijze verwerking	
Memoriseren	Mate waarin studenten bij het studeren gebruik maken van 'van buiten leren'
Analyseren	Mate waarin studenten bij het studeren methodisch te werk gaan
Regulatiestrategieën	
Zelfregulatie	Mate waarin studenten het leerproces zelf willen aansturen
Externe regulatie	Mate waarin studenten op de docent/lector of het leermateriaal vertrouwen om het leerproces aan te sturen
Stuurloos leergedrag	Onduidelijkheid over hoe studenten hun leerproces moeten aansturen

In het licht van de onderzoeksdoelen werden twee longitudinale datasets geanalyseerd. In een eerste steekproef werden studenten gevolgd vanaf het laatste jaar secundair onderwijs tot halverwege het tweede jaar hoger onderwijs. Zoals weergegeven in onderstaande figuur, werd de LEMO vragenlijst op vijf meetmomenten afgenomen. Een tweede steekproef volgde een cohort studenten tijdens hun professionele bacheloropleiding aan een hogeschool. Zij werden bevraagd op drie meetmomenten gedurende de drie jaren in het hoger onderwijs.



Figuur 1: Overzicht van de meetmoment voor de twee steekproeven (SO= secundair onderwijs; HO= hoger onderwijs; MM= meetmoment)

Resultaten

De resultaten van dit proefschrift geven aan dat studenten evolueren in de richting van diep en zelf-gereguleerd leren. Daarnaast blijkt dat deze evolutie in leerstrategieën verschilt doorheen verschillende fases. Tijdens het laatste jaar in het secundair onderwijs blijven leerstrategieën redelijk stabiel. Tijdens de eigenlijke overstap naar het hoger onderwijs daarentegen is er een transitiesprong waarbij alle verwerkings- en regulatiestrategieën toenemen. Tijdens deze transitie stijgt dus ook de mate waarin studenten de leerstof van buiten leren (memoriseren) en het stuurloos leergedrag. Na de transitie, is er een dalende trend in beide leerstrategieën.

Naast het bekijken van hoe studenten gemiddeld evolueren, werd in dit proefschrift ook nagegaan of studenten verschillen in hun groei doorheen de tijd (differentiële verandering). Met andere woorden, volgen de meeste studenten het gemiddelde traject of zijn er studenten die een snellere of tragere verandering doormaken in hun leerstrategieën? Met betrekking tot de eerste steekproef geven de resultaten aan dat studenten voor alle schalen verschillend evolueren doorheen de tijd, behalve voor de schaal memoriseren. Voor de schalen relateren en structureren, kritisch verwerken en analyseren is in deze differentiële groei ook een patroon zichtbaar: studenten die hoger scoren op het eerste meetmoment (november 6 SO) groeien minder sterk doorheen de tijd. Studenten die dus initieel lager scoren op deze drie leerstrategieën, maken een inhaalbeweging.

Vanuit de bevindingen met betrekking tot het tweede onderzoeksdoel - nagaan hoe drie statistische keuzes de parameterschattingen van de verandering beïnvloeden - formuleert dit onderzoek enkele aanbevelingen. Deze aanbevelingen zijn gericht aan onderzoekers die leerstrategieën of andere concepten in kaart brengen aan de hand van zelf-rapportage vragenlijsten met Likert-schalen. Ten eerste, strekt het tot aanbeveling om de items zelf mee te nemen in de analyse in plaats van gebruik te maken van somscores. Op die manier wordt de ruis of error gescheiden van de eigenlijke verandering in leerstrategieën die we wensen in kaart te brengen. Ten tweede, wordt aangeraden om de assumptie van longitudinale meetinvariantie na te gaan en enkel verandering in een leerstrategieschaal te schatten indien deze assumptie niet te sterk wordt geschonden. Met name, wanneer ten minste alle factorladingen gelijk zijn gedurende de meetperiode en de moeilijkheidsgraad van minstens twee items per schaal. Ten derde, tonen de resultaten aan dat de manier waarop met ontbrekende data wordt omgegaan, de parameterschattingen en de conclusies hieruit kan beïnvloeden. Daarom wordt aangeraden om de ontbrekende data mee te nemen in de analyse eerder dan de respondenten met één of meer ontbrekende waarden te verwijderen.

Vervolgonderzoek

In het proefschrift worden verschillende suggesties geformuleerd voor vervolgonderzoek. Zo is het inzicht uitgebreid in *hoe* leerstrategieën evolueren tijdens en na de overstap van het secundair naar het hoger onderwijs. Maar het begrip over *waarom* studenten tijdens de ene periode redelijk stabiel blijven en tijdens een andere periode sterk veranderen, dient verder verkend te worden in vervolgonderzoek. Toekomstig onderzoek kan zich richten op het verklaren van verandering. Bijvoorbeeld, door na te gaan of studenten die de leeromgeving in het hoger onderwijs ervaren als sterk verschillend van de leeromgeving in het secundair onderwijs, een sterkere toename in leerstrategieën rapporteren tijdens de transitieperiode. Ten slotte richt vervolgonderzoek zich bij voorkeur ook op de overstap van hoger onderwijs naar de arbeidsmarkt. Door laatstejaarsstudenten uit het hoger onderwijs op te volgen in hun overstap naar de arbeidsmarkt, kan worden nagegaan worden of deze overstap opnieuw zorgt voor een transitiesprong in leerstrategieën en welke leerstrategieën afgestudeerden het meest nodig hebben om levenslang te leren.

Implicaties voor de onderwijspraktijk

De bevindingen in dit proefschrift hebben eveneens implicaties voor de onderwijspraktijk. De resultaten dit onderzoek beschrijven hoe leerstrategieën veranderen tijdens en na de overstap van het secundair onderwijs naar het hoger onderwijs, maar het is onduidelijk in welke mate de studenten die deelnamen aan dit onderzoek zich bewust waren van deze verandering. Daarom is een belangrijk element in initiatieven rond leerbegeleiding het informeren van studenten over de verandering in hun leerstrategieën en over hoe leerstrategieën gelinkt zijn aan uitval en studiesucces.

Een tweede implicatie voor de praktijk heeft betrekking op de transitiesprong tijdens de eerste maanden in het hoger onderwijs. Diepe verwerking en zelfregulatie nemen toe, maar ook de mate waarin studenten de leerstof memoriseren stijgt, evenals de stuurloos leergedrag. Deze toename is niet van tijdelijke aard. Halverwege het tweede jaar in het hoger onderwijs scoren studenten nog steeds hoger op memoriseren en stuurloos leergedrag dan aan het einde van het zesde jaar secundair onderwijs. Daarom is het van belang om initiatieven rond leerbegeleiding zo vroeg mogelijk in het eerste jaar hoger onderwijs in te bedden. Dit om in te spelen op de veranderlijkheid in leerstrategieën tijdens de eerste maanden in het hoger onderwijs en om tegelijkertijd de minder wenselijke verandering in memoriseren en stuurloos leergedrag tegen te gaan.

Conclusie

Dit proefschrift levert een belangrijke bijdrage aan het onderzoek naar de verandering in leerstrategieën van studenten. Zo werd de verandering in leerstrategieën meer accuraat in kaart gebracht, door rekening te houden met error, door het testen van de assumptie van longitudinale meetinvariantie en door studenten met ontbrekende data mee te nemen in de analyse. De bevindingen over de invloed van de statistische keuzes op de parameterschattingen en de aanbevelingen daarover zijn relevant voor onderzoekers, ook voor zij die longitudinale verandering in kaart brengen in andere concepten dan leerstrategieën. Daarnaast biedt dit proefschrift verdieping op eerder onderzoek door de evolutie in leerstrategieën na te gaan tijdens de overstap van secundair naar hoger onderwijs. Hieruit blijkt dat in bepaalde periodes leerstrategieën redelijk stabiel blijven (bijvoorbeeld het laatste jaar van het secundair onderwijs) en in andere periodes zij meer veranderen (bijvoorbeeld, tijdens de eerste maanden in het hoger onderwijs).