Learning and Individual Differences 62 (2018) 49–61

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Conceptual change and knowledge integration as learning processes in higher education: A latent transition analysis☆

Maja Flaig[a,*], Bianca A. Simonsmeier[a], Anne-Kathrin Mayer[b], Tom Rosman[b], Johannes Gorges[a,b], Michael Schneider[a]

[a] Educational Psychology, University of Trier, Germany
[b] Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID), Trier, Germany

A B S T R A C T

Conceptual change, that is, a restructuring of incompatible prior knowledge, is a well investigated learning mechanism in school children's acquisition of new concepts. An understanding of academic concepts is also a central learning goal of higher education. However, there is almost no research on conceptual change and knowledge integration in higher education. We tracked 137 undergraduate psychology students' concepts of human memory longitudinally over four semesters. A latent profile transition analysis (LPTA) showed that the students' development followed six transition paths between four knowledge profiles. These developmental pathways were well-ordered, indicated a general trend from fragmented knowledge to integrated scientific knowledge, and correlated with the students' university grades and with an additional test of memory understanding. The findings highlight the importance of conceptual change, in particular, knowledge integration in higher education, and exemplify the usefulness of LPTA for modeling individual differences in longitudinal changes of multidimensional knowledge structures.

1. Introduction

Students' understanding of academic concepts, for example, force in Physics, supply and demand in Economy, or human memory in Psychology, is a central learning goal of higher education. Conceptual understanding is a cornerstone of professional expertise (Tynjälä, 1999). It helps learners make predictions, explain observations, reason about the interrelations of facts, infer new knowledge, choose between alternative procedures, and construct new problem solving strategies (Goldstone & Kersten, 2003; Machery, 2010; Rittle-Johnson, Schneider, & Star, 2015). Accordingly, the European Qualification Framework for Lifelong Learning lists “advanced knowledge of a field of work or study, involving a critical understanding of theories and principles” (European Commission, 2008, p. 12) as a central qualification for reaching a Bachelor's degree.

In many cases, conceptual change, that is, a restructuring of the learner's prior knowledge, is a necessary part of acquiring new conceptual knowledge (Carey, 1985; Posner, Strike, Hewson, & Gertzog, 1982; Vosniadou, 2008). Prior knowledge guides and constraints the interpretation of new knowledge and its encoding in memory. It often stems from observations and explanation attempts in everyday life outside formal instruction and thus can be incompatible with the scientific concepts to be learned (Carey, 1992). This explains why acquiring an understanding of academic concepts can be so difficult. Conceptual change is investigated by educational, developmental, cognitive, and philosophical scientists, who found the approach productive in content domains as diverse as physics, chemistry, biology, mathematics, medicine, and the social sciences (M. Schneider, Vamvakoussi, & Van Dooren, 2012; Vosniadou, 2008). Some of the past findings have direct implications for the design of effective learning environments, for example, school instruction (Duit, Treagust, & Widodo, 2008), professional development programs for teachers (e.g. Hewson, Tabachnick, Zeichner, & Lemberger, 1999), and projects to foster instructional quality in schools (Beeth et al., 2003).

Empirical research on conceptual change of students in higher education is virtually non-existent, despite the importance of academic concepts as learning goals in higher education. Almost all studies on conceptual change focused on students in K-12 schools or even younger children. In the present study, we examined to what extent conceptual change is still a relevant learning mechanism in higher education and whether it leads to similar developmental patterns in higher education as it does in K-12 school learning. We expected conceptual change to

☆ This work was supported by the German Joint Initiative for Research and Innovation with a grant acquired in the Leibniz Competitions 2013 (grant number SAW-2013-ZPID-1 195).
* Corresponding author at: Educational Psychology, University of Trier – Division I, 54286 Trier, Germany.
E-mail address: flaig@uni-trier.de (M. Flaig).

https://doi.org/10.1016/j.lindif.2017.12.008
Received 1 December 2016; Received in revised form 27 November 2017; Accepted 9 December 2017

still be relevant in higher education, because learning by conceptual change has been described as a general human learning mechanism that is relevant independently of age group and content domain. For example, there are some similarities between children's thinking processes when acquiring new concepts and scientists' thinking processes when developing a new theory, which hints at an age-general importance of conceptual change (Gentner et al., 1997).

1.1. Knowledge fragmentation and integration

Research with school-aged children found that the fragmentation and integration of knowledge are two central component processes of conceptual change (e.g., M. Schneider & Hardy, 2013). Networks of conceptual knowledge in long-term memory can comprise types of elements, such as observations, hypotheses, explanations, analog mental models, mental images, category exemplars, and subjective theories (Goldstone & Kersten, 2003; Machery, 2010). These elements have been acquired in situation as diverse as conversations with peers, everyday-life observations, internet videos, school instruction, books, or movies. Learners do not always understand how these different and sometimes even conflicting pieces of knowledge relate to each other, and store them independently in long-term memory. This leads to fragmented knowledge.

Another source of knowledge fragmentation is the fact that storing correct concepts in long-term memory does not necessarily erase related misconceptions. Converging evidence from reaction times studies with sentence verification tasks (Potvin, Masson, Lafortune, & Cyr, 2015; Shtulman & Valcarcel, 2012), multiple choice tests (M. Schneider & Hardy, 2013), and interviews (diSessa, Gillespie, & Esterly, 2004) shows that naïve misconceptions and scientifically correct concepts or parts thereof can co-exist in long-term memory and do so frequently, not only in children, but also over the life-span (Shtulman & Harrington, 2016). Pieces of fragmented knowledge are activated dependent on the context (diSessa et al., 2004), so that learners do not necessarily realize when they hold pieces of knowledge in long-term memory that support or contradict each other.

Thus, the integration of pieces of knowledge into a coherent overarching knowledge structure is an important aim of instruction (Linn, 2006; M. Schneider, 2012). This can include connecting previously isolated pieces of knowledge in memory and subsuming previously unrelated concepts under a general principle. Understanding these relations can help students to differentiate better between normatively correct and incorrect ideas, thus, leading to more homogeneous and more correct knowledge. Teachers can stimulate knowledge integration by eliciting students' knowledge, adding new normative concepts, helping students to develop criteria for evaluating alternative conceptions, and by encouraging students to compare the alternatives and to sort out inadequate conceptions (Linn, 2006).

1.2. A latent profile transition analysis of fragmented and integrated knowledge

Developmental patterns of co-existing pieces of knowledge can be investigated by latent profile transition analyses (LPTA), as demonstrated by three studies with school children on knowledge development in mathematics and science (Kainulainen, McMullen, & Lehtinen, 2017; McMullen, Laakkonen, Hannula-Sormunen, & Lehtinen, 2015; M. Schneider & Hardy, 2013). M. Schneider and Hardy (2013) investigated third-graders' understanding of floating and sinking of objects in liquids. The children participated in several sessions of either (1) a constructivist learning environment with a high degree of instructional support given by the teacher, or (2) a constructivist learning environment with a low degree of instructional support, or (3) a baseline control group without instruction on the topic (see Hardy, Jonen, Möller, & Stern, 2006, for details of the interventions). Before and after the instruction as well as one year later the children indicated in a multiple choice test how strongly they agreed with a number of statements. Each statement described (a) a common misconception, (b) an everyday life explanation, which has some explanatory power in everyday life but does not hold up to systematic scientific evaluation, or (c) the relevant scientifically correct concepts. The three sum scores indicating how often each child agreed with misconceptions, everyday conceptions, or scientific concepts were used as indicators of latent profile memberships at the three measurement points in a latent profile transition model. The parameters of this model were estimated so that the similarity of the scores of persons grouped in the same latent profiles was maximized and the similarity of the scores of persons grouped in different latent profiles was minimized (see Hickendorff, Edelsbrunner, McMullen, Schneider, & Trezise, this issue, for methodological details). Thus, the latent profiles indicated groups of learners who had the same configuration of misconceptions, everyday conceptions, and scientific concepts. In addition to the profile characteristics and memberships at each measurement point, the model estimation also yielded the frequencies of the profile transitions over time. These were interpreted as pathways of conceptual change.

The model results in the study by M. Schneider and Hardy (2013) indicated five latent classes. Some of these had mean score profiles that indicated integrated knowledge, that is, high profile mean scores for misconceptions only or for scientific concepts only. Other profiles indicated fragmented knowledge. In these profiles, two or three of the scores for misconceptions, everyday concepts, or scientific concepts were higher than the sample mean. Most participants were on one of seven developmental pathways between these five profiles over time. These transition paths indicated a general trend from misconceptions and fragmented knowledge towards more correct and integrated knowledge. However, there were strong individual differences. Knowledge fragmentation increased on some paths and decreased on others. About 20% of the children still had fragmented knowledge, that is, co-existing misconceptions, everyday concepts, and scientific concepts, even one year after participating in the constructivist learning environment. The instructional condition was related to the frequency of the transition paths. The constructivist learning environment with a high degree of instructional support led to the most integrated knowledge and the untreated control group to the least.

The other two studies using latent transition analyses to investigate knowledge development traced school students' knowledge of rational numbers over time (Kainulainen et al., 2017; McMullen et al., 2015). Similar to M. Schneider and Hardy's findings, students' knowledge of rational numbers was captured by a small number of knowledge profiles and systematic transition paths between these profiles over time, some of which could be interpreted in terms of conceptual change. However, the generalizability of these findings to other age groups and content domains remains unclear (cf. Edelsbrunner, Schalk, Schumacher, & Stern, this issue; McMullen, Van Hoof, Degrande, Verschaffel, & Van Dooren, this issue).

1.3. Is conceptual change still relevant in higher education?

In the current study, we used latent profile transition analysis to investigate whether conceptual change and, more specifically, knowledge fragmentation and integration still can be found in higher education and whether their quality is related to the learning outcomes, that is, students' grades. There are at least three reasons to expect that this might not be the case. First, students in higher education successfully participated in school instruction, perhaps leading to a firm fundament of correct and integrated knowledge that higher education can build on. Second, students in higher education tend to have better developed meta-cognitive strategies than school children (Weil et al., 2013). Thus, they might be able to monitor, identify, and understand the confirmatory or contradictory relations between their prior knowledge and the knowledge to be learned making knowledge integration a trivial process. Finally, arguably, the content of higher education programs is more abstract than the content of school lessons, so that fewer interferences between prior knowledge from everyday life and the new knowledge may occur.

On the other side, there are also a number of reasons why conceptual change could still affect student learning in higher education. First, the structure and the functions of human memory do not change fundamentally between K-12 school and higher education. Adults' conceptual knowledge is still organized as a network, so restructuring this network might be necessary when prior knowledge and new knowledge are incompatible. Second, empirical research shows that university students still have many misconceptions, which are stable, hamper subsequent learning, and require restructuring (Merz, Dietsch, & Schneider, 2016; Shtulman & Valcarcel, 2012). Third, a recent meta-analysis on undergraduate science course innovations found that so-called conceptually-oriented learning tasks had a substantial effect on student achievement. Averaged over nine studies the effect size was $d = 0.47$ ($SD = 0.70$). The authors defined that tasks are conceptually oriented when they "elicit students' level of understanding of key science concepts, […] engage students in conceptual schemes within a topic rather than isolated facts, […] and engage students with real-world problems in creative ways that reflect a conceptually integrated understanding of the content" (Ruiz-Primo, Briggs, Iverson, Talbot, & Shepard, 2011, p. 1269). Based on these findings we hypothesize that our latent profile transition analysis will find evidence for conceptual change and, in particular, for knowledge fragmentation and integration, in higher education.

1.4. Psychology students' concepts of human memory

We tested our hypotheses in a longitudinal study on Psychology students' understanding of human memory. Specifically, we assessed to what extent the students had the misconception of memory as a place for the static storage of information and to what extent they had the correct concept of memory as a dynamic system involving the construction and re-construction of knowledge at four measurement points in the first four semesters of the Bachelor program in Psychology (Lynn & McConkey, 1998). Examples of memory processes that modify the information to be stored are interference (Bjork, 1992), chunking (Gobet et al., 2001), and source monitoring (Johnson, Hashtroudi, & Lindsay, 1993).

Human memory is a complex causal system and one of the central constructs of Psychology (Baddeley, Eysenck, & Anderson, 2009). The search term *memory* has > 190.000 hits in the literature database PsycInfo (November 2016). These articles are from fields as diverse as cognitive, educational, developmental, clinical, social, forensic, and biological psychology. An understanding of memory properties and its processes is also important in numerous psychological professions, and a flawed understanding of human memory in professionals who work in the clinical or legal context can have negative consequences (Garry, Loftus, & Brown, 1994). As far as we know, there is no standardized test of students' concepts of human memory. Therefore, we used a self-developed test in our study.

To assess the criterion validity of our new test, we additionally presented the participants with the Implicit Memory Theory Scale (IMTS; Niedźwieńska, Neckar, & Baran, 2007), which tests how skeptic individuals are with respect to the credibility of autobiographical memory. We expect that students who have a concept of human memory as a dynamic system that constructs and re-constructs information will be more skeptical with respect to the validity of autobiographic memories than other students.

1.5. The current study

In sum, conceptual understanding is a central learning goal of higher education. Yet, there is almost no published research on conceptual change in higher education students. For this reason, we used a latent profile transition model with longitudinal data from university students in the current study. We constructed the measures so that the latent profiles can be interpreted in terms of knowledge fragmentation and integration and that any profile transition can be interpreted as signs of conceptual change. We tried to replicate M. Schneider and Hardy's (2013) findings on a general conceptual level, in particular, the trend from misconceptions to scientific concepts, the trend from fragmented to integrated knowledge, and the persistence of fragmented knowledge in some learners even after instruction.

Based on the previous findings of studies with school children we had the following hypotheses for our study in higher education: (1) Persons differ in their knowledge profiles and transition paths. However, the number of profiles and paths is relatively small. (2a) There are transitions between latent profiles differing in their knowledge profiles over time, giving evidence of conceptual change in higher education. (2b) The transitions reflect an overall developmental trend from profiles with higher scores of misconceptions towards profiles with higher scores of scientific concepts. (3a) At least one of the knowledge profiles indicates fragmented knowledge by high agreement with mutually incompatible concepts. (3b) The profiles indicating fragmented knowledge will become less frequent but will not disappear completely over time. (4) The latent profiles differ in their mean scores on the Implicit Memory Theory Scale indicating that students' profiles of knowledge of human memory are related to how much trust they put in autobiographic memory. (5) The latent profiles differ in their grades on the human memory course showing that students' profiles of knowledge of human memory are related to grades in higher education.

2. Method

2.1. Participants

137 students enrolled in a Bachelor of Psychology program at a mid-sized university in a mid-sized German city participated at T1 in our study. Of these, $N = 126$ participated again at T2, $N = 116$ at T3, and $N = 115$ at T4. Almost all participants were German, and all were fluent speakers of German. At T1, the sample mean age was 20.4 (min = 18; max = 31) and all students were at the beginning of their first semester in the program. About 82% of participants in the sample were female. This proportion is slightly higher than the proportion of all females in the program, which was 60–70%. Participation in the study was anonymous and voluntary. To keep dropout at a minimum, participants were financially compensated with €25 per wave of the longitudinal study and an additional completion bonus of €100 if they had participated in all waves.

The study was conducted in full accordance with the Declaration of Helsinki and the APA Ethics Code (American Psychological Association, 2002). Prior to their participation, students received an informed consent form including, among others, the following information (based on the APA Ethics code): (1) a statement on the purpose of the research as well as the expected study duration and procedures, (2) a statement that participation is voluntarily and that it may be terminated at any point; (3) a statement that there are no potential risks, discomfort or adverse effects with regard to their participation; (4) a statement that data is collected anonymously, and that even though some personal data (e.g., e-mail addresses) will be collected for organizational purposes, these data will not be used to identify individual participants and will be deleted as soon as possible.

2.2. Procedure

The students were tested longitudinally at four measurement points covering the first four semesters of the Bachelor in Psychology program. Data collection took place between November 2013 and May 2015. Baseline data collection (T1) took place during the first six weeks of the participants' first semester in the program, followed by three consecutive waves of measurement (T2, T3, T4) at the beginning of the second, third, and fourth semester, respectively. Each wave consisted of two parts: a home module and a subsequent lab module. The home modules included several self-report measures and were completed online under unsupervised conditions before the respective lab sessions took place; answering the questionnaires took between 30 and 50 min (dependent on the wave). In the lab modules, knowledge and achievement tests were conducted in the university's computer labs. Groups of 1 to 25 participants were supervised by student experimenters and completed the tests individually and at their own pace. Each lab module took about 120 min.

2.3. Measures

2.3.1. Concepts of human memory

Conceptual understanding of human memory was assessed by a self-constructed multiple choice test at the beginning of each of the four lab modules. Previous studies (cf. M. Schneider & Hardy, 2013) have found that students' answers to multiple choice items are well in line with their answers to interview questions that aim to assess the same concepts. The tasks were presented on a computer screen and the answers were entered by mouse clicks. Each of the nine test tasks began with a written description of a situation in which a certain memory process is particularly important. These situations were derived from classical experiments in memory research (Brooks, 1967; Chase & Simon, 1973; Ericsson, Chase, & Faloon, 1980; Goff & Roediger, 1998; Henik & Tzelgov, 1982; Hovland & Weiss, 1951; Loftus, 1975; Melcher & Schooler, 1996; W. Schneider, Körkel, & Weinert, 1989). In three situations, interference between elements in memory (Bjork, 1992) can be expected, in three tasks the chunking of elements into bundles of information (Gobet et al., 2001; Johnson et al., 1993), and in three tasks source monitoring mechanisms (Johnson et al., 1993). Each of the nine publications was cited between 181 and 3286 times, had been replicated repeatedly, and was (still) widely accepted by the scientific community.

In each of the nine tasks, the learners were presented with six statements about memory processes that might be relevant in this situation (see the example task in Fig. 1 and Appendix B for examples of all three types of tasks). There were three types of statements: (a) *Correct processing statements* described the memory process actually occurring in that situation as firmly established by empirical research, as described in the previous paragraph. These statements were compatible with the view of memory as a dynamic system involving the construction of knowledge. (b) *Incorrect processing statements* also described dynamic knowledge construction processes in memory, however, only processes not occurring or being irrelevant in that specific situation as shown by established empirical research. Incorrect processing statements were partly correct (because they describe memory as a dynamic system) and partly incorrect (because they described processes not relevant in the respective situation). (c) *Static storage statements* described memory in that situation as a place for the static storage of information, that leaves the nature and the content of the stored information unchanged. They did not conform to the established empirical findings and the view of the scientific community and can be seen as misconceptions.

The participants evaluated each statement on a seven-point rating scale from definitely incorrect to definitely correct. In each task, only one memory process was relevant. Therefore, there was one correct processing statement, but three incorrect processing statements, and two static storage statements for each task. Half of the statements used a negative wording, e.g., "The two groups do not differ". The order of the nine tasks and of the six statements in each task was randomized separately for each person. We reversed the scores for the incorrect statements and then computed a separate sum score for static correct processing statements, incorrect processing statements, and static storage statements, respectively.

The test was not tailored to the content of the specific course. Human memory is referred to in many courses of the Bachelor Psychology program. We developed the test to reflect students' accumulating knowledge of memory over several semesters and courses. Still the test content is more similar to the lecture "Human Memory" than to any other course in the students' program. Students typically attend this lecture during their first semester in the program.

2.3.2. Implicit Memory Theory Scale

At T3, we administered a German translation of the IMTS (Niedźwieńska et al., 2007) as part of the at-home module. One researcher from our lab translated it from English into German. To check for the validity of the translation, a second researcher from our lab translated the German items back into English without knowing the original English items. In case of disagreements between the two English versions, the German items were modified by both researchers together. The IMTS is a standardized instrument to test how individuals judge the credibility of autobiographical memory. Low skepticism is associated with misconceptions of human memory, such as the belief that memory is a static storage and that recollection is an accurate representation of real events, whereas high skepticism is associated with the belief that memories are reconstructed during recall and that memories are prone to qualitative changes. The global original scale has good psychometric properties (Niedźwieńska et al., 2007). In the original study, the internal consistency was good with a Cronbach's alpha of 0.83 and the retest reliability was also good, with a value of 0.81 for a two-week interval and 0.74 for a seven-month interval. The IMTS demonstrated to be sensitive to knowledge differences between psychologists and non-psychologists, as well as to knowledge differences between psychology students before and after instruction on autobiographical memory. We report the mean and standard deviation of the IMTS scores in our sample in Table 1.

2.3.3. Grades

At T4, the university administration sent us the grades for all the exams the participants in our study had completed so far. In the German grading system, grades range from 1 to 5, with smaller numbers indicating better performance. The students are relatively free in choosing the sequence of their exams, so the types and numbers of completed exams varied between participants. For our analysis, we only used the grades of a course in general psychology, which covers the topics *human memory*, *learning*, *motivation* and *emotion*, and thus, is the best proxy for human memory competence. At T4, 82% of the sample had completed the respective exam. The mean and standard deviation of the grades is reported in Table 1.

2.4. Statistical analysis

In the latent profile transition model, we specified a latent profile variable for each of the four measurement points. Each latent profile variable had the sum scores for correct processing, incorrect processing, and static storage at the respective measurement point as indicators of the persons' profile memberships. The profile means of the indicator variables were constrained to be equal over time, so that fewer parameters had to be estimated and that the profiles had the same interpretation at all four measurement points. The number of participants in each profile and the variance of the indicator variables was allowed to differ between measurement points. The profile membership at each time point was used as a predictor of the profile membership at the respective following time point, so that intercepts and regression weights of three multinomial logistic regressions were estimated. MPlus uses these parameters to compute transition probabilities and transition paths between the latent profiles over time in the model. Overall, the model has 63 free parameters. These are the 12 means and 12 variances of the profile indicators (i.e. for the three indicators at the four time points), 12 regression intercepts (i.e. for four latent profiles minus one

Wild horses

Two groups of children participate in the study.
The children in one of these groups (Group A) are twelve year-olds, who know little about horses in general.
The children in the other group (Group B) are eight year-olds, who know a lot about horses in general.
The two groups do not differ in terms of the children's intelligence and the number of boys and girls in the group.
All participants read a simple text about wild horses that teaches them what kinds of species exist, where they are found and what they need to live. Afterwards, the children are asked to answer questions assessing their knowledge and comprehension, based on the text from memory.

How sure are you that each of the following statements is correct or incorrect?

	Definitely incorrect … Definitely correct
Group A will perform markedly better than Group B because they have four more years of practice in reading and remembering from texts than the younger children.	O O O O O O O
Group B will perform markedly worse than Group A, because the children are on a lower stage of cognitive development and therefore cannot process information so well.	O O O O O O O
Group B will perform markedly better than Group A because they have more prior knowledge and therefore can store information from the text in a more structured way.	O O O O O O O
Group A will perform markedly worse than Group B because the memory of twelve year-olds is partly impaired already due to hormonal changes during puberty.	O O O O O O O
Both groups will perform almost equally well because they have read the same text with the result that each person has memorized the same information.	O O O O O O O
There barely will be any difference between the two groups because their intelligence is similar and therefore they can process information almost equally well.	O O O O O O O

Fig. 1. Example task in the test of concepts of human memory (in this figure translated from German to English). The correct answer is in bold print here.

Table 1
Descriptive statistics of all unstandardized scores.

	M	*SD*	*N*
Static storage T1	1.187	0.609	137
Incorrect processing T1	2.004	0.599	137
Correct processing T1	3.795	0.715	137
Static storage T2	0.798	0.591	126
Incorrect processing T2	1.664	0.654	126
Correct processing T2	4.414	0.738	126
Static storage T3	0.736	0.628	116
Incorrect processing T3	1.488	0.693	116
Correct processing T3	4.410	0.823	116
Static storage T4	0.584	0.579	114
Incorrect processing T4	1.319	0.714	114
Correct processing T4	4.490	0.859	114
IMTS T3	30.997	10.171	116
Grades T4	1.994	0.897	111

at four points in time), and 27 regression weights (for regressions of four profiles minus one at one time point to four profiles minus one at the respective next time point times three pairs of time points).

The data were prepared for structural equation modeling using SPSS IBM® Version 23. The latent profile transition model was analyzed in MPlus 7.31 (Muthén & Muthén, 1998–2016). We used the maximum likelihood estimator with robust standard errors (MLR), which is the default estimator for latent transition models in MPlus. The model was estimated using the MPlus default handling of missing data according to which all available data is used to estimate the model (Muthén & Muthén, 1998–2016). When there was missing data, the covariance coverage of the variables used in our analysis was between 0.825 and 0.920 which far exceeded the minimum covariance coverage of 0.100. In latent profile, latent class, and latent transition analysis, it is important to make sure that the best log likelihood value is replicated several times and to avoid improper solutions due to local maxima. We therefore increased the default number of random starts in the initial stage to 400 sets and the number of final stage optimizations to 100. We used unstandardized (raw) scores of the indicator variables for the model estimation to yield unbiased model results. These scores are reported in Table 1. After the model estimation and to aid the interpretation of the results, we standardized the profile mean values and standard deviations to T-scores, which have a mean of 50 and a standard deviation of 10. This was done by first subtracting the sample mean from the profile mean and dividing the result by the sample standard deviation separately for each profile indicator in each profile. The sample means and *SD*s were $M = 0.86$, $SD = 0.48$, for static storage, $M = 1.66$, $SD = 0.59$, for incorrect processing, and $M = 4.22$, $SD = 0.65$, for correct processing. Following, the resulting values were multiplied by 10 and added to 50. We report these standardized scores in the following to aid interpretation.

3. Results

In the following three subsections, we first describe how we determined the number of latent profiles, then characterize the latent profiles, and then describe the transition paths between the latent profiles over time. Finally, we analyze the relationship between the latent profiles and outside criteria, that is, the IMTS and course grades.

3.1. Determining the number of latent profiles

We selected the number of latent profiles based on the recommendations given by Nylund, Asparouhov, and Muthén (2007). The best fitting model is determined by repeatedly estimating the model with increasing numbers of classes or profiles and comparing their model fits. There is no definite standard for deciding on the number of latent profiles (Nylund et al., 2007). Better-fitting mixture models are characterized by lower comparative fit indices, i.e., lower values in the

Table 2
Fit indices for the latent transitions models with one to five latent profiles.

Index	One profile	Two profiles	Three profiles	Four profiles	Five profiles
Log likelihood	− 1513	− 1350	− 1278	− 1245	− 1212
Free parameters	24	25	41	63	91
AIC	3074	2749	2638	2617	2607
BIC	3145	2822	2758	2801	2873
Sample-size adjusted BIC	3069	2743	2628	2601	2585
Entropy	–	0.840	0.832	0.860	0.841

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion.

Akaike information criterion (AIC), Bayesian information criterion (BIC), and sample-size adjusted BIC. Another way to examine model fit is to compare the classification quality of individuals into latent profiles between models that differ in the number of profiles assumed. Classification quality or *entropy* is measured on a scale that ranges between zero and one, with a value of one indicating perfect classification of individuals into latent profiles (Clark & Muthén, 2009).

We repeatedly estimated the model, each time with a different number of specified latent profiles. For each model, the best log likelihood value was replicated several times with the initial specification of 400 sets of random starts and 100 iterations, that is, 99 times in the one-class model, 100 times in the two-class and three-class models, 25 times in the four-class model, and 8 times in the five-class model. In the second step, in which we ran the models again with twice the number of random starts and final stage iterations, each of these values was replicated again. Table 2 presents the model fit indices for the models with one to five latent profiles. The models with two to five latent profiles had entropies above 0.80, which is high according to Clark and Muthén (2009). The BIC was lowest for the three-profile model, whereas the AIC and sample size-adjusted BIC were lowest for the five-profile model, and entropy was highest in the four-profile model. Therefore all three models fit the data approximately equally well. We chose the four-profile model, because the four-profile model but not the three-profile model included a profile indicating strong misconceptions (high values for static processing only). Even though only small sample proportions showed this profile in our study, it is still interesting, because it corresponds to the lowest level of knowledge of human memory, thus is the starting point of competence development in a domain, and had also been found in previous research (cf. M. Schneider & Hardy, 2013). Furthermore, we chose the four-profile model over the five-profile model because it was more parsimonious and had a lower BIC. All subsequent analyses were conducted with the four-profile model.

3.1.1. Interpretation of the profile means

Table 3 lists the T-standardized indicator means for the four latent profiles. In Fig. 2, the standardized scores are represented graphically. To aid interpretation, we tested for static storage statements, incorrect processing statements, and correct processing statements in each latent profile at each measurement point whether the latent profile mean significantly deviated from the overall sample mean. For example, we fixed the mean of static storage in profile 1 at the overall sample mean of static storage while fixing the other means at the values of the initial 4×4 latent transition model. In total, we computed $3 \times 4 = 12$ models for the three mean scores across the four latent profiles. Then, we computed log likelihood ratio chi-square difference tests to test for the difference between the fixed mean models and the initial model. We computed the robust test statistic for MLR estimation (Satorra & Bentler, 2010) as reported by Asparouhov and Muthén (2010; formula 2) and adjusted the significance level for the number of comparisons using the Bonferroni method, yielding a significance level of $\alpha = \frac{.05}{12} = 0.004$. Table 3 shows the results of these comparisons.

We interpreted the latent profiles based on the standardized mean scores and the results of the computed log likelihood ratio chi-square difference tests, and we assigned labels to them (see column 2 in Table 3). We labeled the profile C1 *misconceptions profile*, because students with this profile show the highest scores for the static storage misconceptions of all classes, average scores for incorrect processing statements, and below-average scores for correct processing statements, even though the first and last deviation were not significant in the chi square difference tests. However, it should be noted that only a small number of individuals were assigned to the misconceptions profile, yielding large standard errors, and the Bonferroni correction is a rather conservative correction of the significance level. We called the profile C2, which shows significant above-average means on two of the three scales, *fragmented profile* because these learners express inconsistent knowledge: on the one hand they believed that human memory was a static storage and on the other hand they believed that memory processes and re-constructs information. This profile supports our hypothesis H3a, that some learners have fragmented knowledge rather than integrated knowledge. The third profile, C3, indicated that the participants mostly chose the answer category in the middle of the rating scale with average scores on all three scales. We labeled it *indecisive profile*. Finally, we labeled the profile C4 *scientific profile* because it displayed significant above-average means of scientific concepts and significant below-average scores of static storage statements and incorrect processing statements, representing the ideal learning outcome.

To further aid interpretation, we tested whether the latent profiles differed significantly in their means on the three indicator variables. We computed a separate model for each pair of the four latent profiles. In each model, we constrained two class profiles to be equal whereas all other model parameters were fixed to the values found with the original (unconstrained) four-profile model. We corrected for multiple testing using the Bonferroni method and set the significance level at $\alpha = \frac{.05}{6} = 0.008$. Five of the six constrained models had a significantly worse fit than the original four-profile model, as found by likelihood ratio chi-square difference tests, all $ps \leq 0.001$. Only the comparison between the unconstrained model and the model in which the first and the third profile were constrained equal was not statistically significant, $p = 0.012$. The reasons for this comparison missing the critical significance level was the low frequency of the misconceptions profile and

Table 3
Assigned labels for the latent knowledge profiles, sample proportions, standardized scores of the scales' means, and significance of the deviation of each mean from 50.

	Label	Sample proportion in %				Static storage		Incorrect processing		Correct processing	
		T1	T2	T3	T4	*M*	*p*	*M*	*p*	*M*	*p*
C1	Misconceptions profile	7	5	6	4	92	.053	49	1.00	30	.544
C2	Fragmented profile	56	22	17	17	71	< .001*	71	< .001*	39	.028
C3	Indecisive profile	37	46	43	38	40	.281	48	.927	52	.998
C4	Scientific profile	0	27	33	42	26	< .001*	25	< .001*	70	< .001*
	Total	100	100	100	100						

* Significant on the level of $\alpha < 0.004$, corrected for multiple testing.

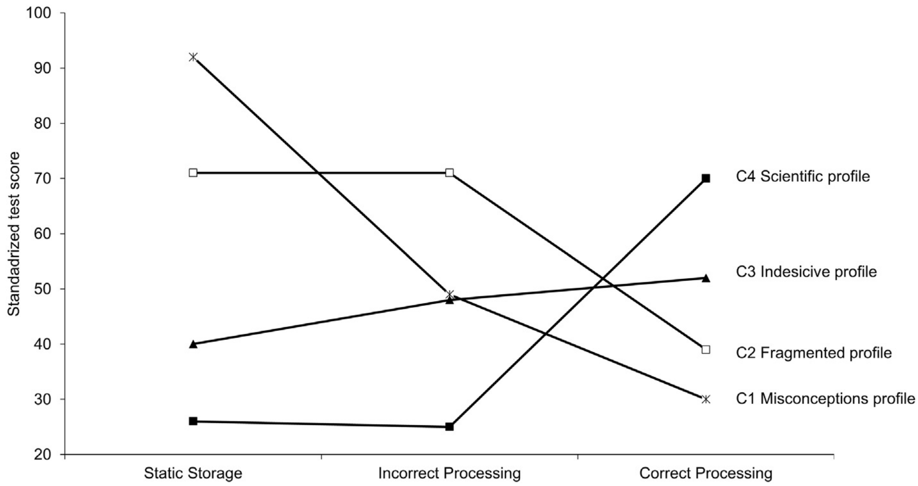

Fig. 2. Diagram of the four knowledge profiles.

the very conservative Bonferroni-corrected critical significance level. Descriptively, as shown in Fig. 2, there is a large difference between these two profiles. Thus, the results indicate that all four class profiles are mutually different and support our first hypothesis, that there are latent profiles of learners differing systematically in their conceptual knowledge (H1).

3.2. Changing profile frequencies on the sample level over time

The proportions of the latent profiles changed substantially from T1 to T4 (see in Table 3), except for the misconceptions profile, which was shown only by a small proportion of the sample (4% to 7%) at all four measurement points. In line with hypothesis H3b, the frequency of the fragmented profile decreased from 56% to 17%, but did not approach 0, thus indicating that some Psychology students still had fragmented knowledge about the memory concept at the end of their fourth semester. The proportion of the indecisive profile increased from T1 to T2 and stayed stable to T3, before finally dropping back to the initial level. The frequency of the scientific profile increased sharply from 0% at T1 to 37% at T2 and remained high afterwards. These findings are in line with our second hypothesis (H2b), that there is an overall trend towards a scientific understanding of human memory.

3.2.1. Strengths of the predictive relations between knowledge profiles

In order to test for possible predictive relationships between the knowledge profiles of T1 to T4, we generated three frequency tables (T1–T2, T2–T3, and T3–T4). We used the profile proportions and the transition probabilities based on the estimated model to compute profile frequencies. Then, we build cross tabulations, including the profile frequencies at one point in time (rows) to predict the profile frequencies at the next point in time (columns). For the predictive value of T1 for T2, the chi-square test indicated a highly significant and strong positive relation, $\chi^2 = 174.37$, $df = 6$, $p < 0.001$, Cramer's $V = 0.798$. The relations were about equally strong for the other points of time, $\chi^2 = 313.32$, $df = 9$, $p < 0.001$, Cramer's $V = 0.876$ (T2 × T3), and $\chi^2 = 266.08$, $df = 9$, $p < 0.001$, Cramer's $V = 0.808$ (T3 × T4). Thus, the knowledge profiles allow predictions about students' future knowledge profiles at a later point in time. Only 23% of the participants did not change their knowledge profile over the course of the study. For the remaining 77% of the sample, information about the profile at one point in time helped to predict changes to another profile. In the following section we take a closer look at the participants' individual transition paths.

3.2.2. Latent transition paths

MPlus reports overall sample latent transition probabilities based on the estimated model across time (for the latent transition probabilities in the current study, see Appendix A). These probabilities are used to calculate latent transition paths. In a model with four profiles at four measurement points, there are $4^4 = 256$ possible latent transition paths between the profiles across time. In line with our Hypothesis 1, the participants were on few of these possible paths. There were only six paths taken by at least 5% of the sample (see Fig. 3), and 81% of the sample was on these six paths. Approximately 17% of the sample was on another eight paths which were taken by at least one person of the sample. The remaining 3% were on paths that were estimated to be taken by less than one person of the sample by MPlus. The fact that a high proportion of the sample followed a small number of developmental paths shows that knowledge development was highly systematic in our sample.

We interpreted the transition paths based on the included profiles and assigned labels to them (see Table 4). The first path, P1, was labeled *increasing correct knowledge* because participants on this path displayed an indecisive profile at T1 and transitioned to the scientific profile at T2, where they stayed throughout the course of the study. The second path, P2, was called *decreasing fragmentation* because individuals on this path transitioned from the fragmented profile to the indecisive profile which reflects a development towards more integrated pieces of knowledge. Path P3 was called *enduring fragmentation*, as participants on this path stayed with the fragmented profile throughout the whole study. Similarly, participants on P4 stayed with the indecisive profile, thus we labeled the path *enduring indecisiveness*. The paths P5 and P6

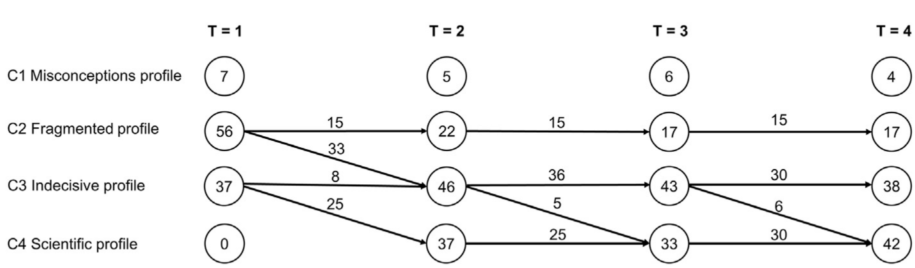

Fig. 3. Diagram of the six transition paths taken at least by 5% of the sample. All numbers are percentages of the sample. The numbers in circles refer to the whole sample (100% in total at each wave). The numbers next to the arrows refer to the proportion of the sample on the most common six transitions paths (81% of the sample in total).

Table 4
Pathways of conceptual change (i.e. model-estimated latent transition paths) in the sample.

Path	Label	Knowledge profile				Proportion of the sample (%)	
		T1	T2	T3	T4	Alone	Accumulated
P1	Increasing correct knowledge	C3	C4	C4	C4	25	25
P2	Decreasing fragmentation	C2	C3	C3	C3	22	47
P3	Enduring fragmentation	C2	C2	C2	C2	15	62
P4	Enduring indecisiveness	C3	C3	C3	C3	8	70
P5	Slowly evolving scientific concepts	C2	C3	C3	C4	6	76
P6	Quickly evolving scientific concepts	C2	C3	C4	C4	5	81
P7-P14	Various paths found for at least one person	Various profiles	Various profiles	Various profiles	Various profiles	17	97
P15-P256	Various paths not found for at least one person	Various profiles	Various profiles	Various profiles	Various profiles	3	100

Note. C1 = misconceptions profile; C2 = fragmented profile; C3 = indecisive profile; C4 = scientific profile.

both included two transitions, and students on these paths started with the fragmented profile, transitioned to the indecisive profile, and finally moved to the scientific profile. Students on P5, transitioned to the scientific profile between T2 and T3, whereas students on P6 transitioned to the scientific profile between T3 and T4. We labeled the paths *slowly evolving scientific concepts* and *quickly evolving scientific concepts* because these transition represent a gradual decrease in the static storage misconception while processing statements are more and more endorsed.

As can be seen in Fig. 3, on the six most frequent transition paths, the participants either stayed in their respective latent profile over time or transitioned from lower-ranking profiles (e.g., the indecisive profile) to higher-ranking profiles (e.g., the scientific profile). There were no transitions in the opposite direction. Between T1 and T2, the majority of the sample (58%) transitioned from a lower-ranking to a higher-ranking profile. Conversely, between T2, T3, and T4 the majority of the sample (72%) stayed in their respective latent profile. This mirrors the fact that most participants successfully attended a lecture on human memory during the first semester of their program.

3.3. Associations without outside criteria

To test for relations between the knowledge profiles and outside criteria, we exported the list of participants' most likely profile memberships to SPSS. To test statistically for associations between profile memberships at T3 and T4, IMTS scores, and grades, IMTS and grades were recoded, so that higher scores indicated higher achievement or better grades. As the number of students differed strongly between the four knowledge profiles, we used non-parametric tests to conduct the analyses. Parametric tests, such as the analysis of variance (ANOVA) have been shown to be robust against some violations of assumptions (Schmider, Ziegler, Danay, Beyer, & Bühner, 2010). However, multiple problems and unequal group sizes lead to serious constraints to the robustness of parametric tests (Lix & Keselman, 1998). As a non-parametric alternative to one-way ANOVA we used the Kruskal-Wallis *H*-test. To follow up on the findings, we used the Jonckheere-Terpstra test (Jonckheere, 1954; Terpstra, 1952), which allows to test whether the medians of the groups are ordered in a meaningful way. In SPSS, the Jonckheere-Terpstra test investigates whether the medians ascend or descend in a pre-defined order. We specified the independent variable in the hypothesized developmental rank order of knowledge profiles, that is, from a misconceptions profile, over fragmented and indecisive profiles to a scientific profile. Table 5 shows the recoded IMTS scores and grades by most likely profile memberships.

As hypothesized (H4), the latent profiles at T3 differed significantly in their IMTS scores as indicated by the Kruskal-Wallis *H*-test, $H(3) = 25.845$, $p < 0.001$. The medians were not in perfect accordance with the assumed developmental rank order, as the median of the fragmented profile C2 was lower than the median of the misconceptions profile C1. However, Jonckheere's test was significant, $J = 3112.500$, $z = 4.546$, $r = 0.422$, indicating that participants' knowledge profiles predicted their achievement on the IMTS in a developmental rank order from the fragmented profile to the scientific profile. Thus, persons with a better understanding of memory in general are also more skeptical about the credibility of autobiographical memories. This is in line with a view of human memory as involving the active processing and reconstruction of information.

Table 5
Descriptive statistics for IMTS scores and grades by most likely profile memberships.

	Label	IMTS T3			Grade T4		
		Mdn	*SD*	*n*	*Mdn*	*SD*	*n*
C1	Misconceptions profile	26.806	10.382	7	2.000	0.628	5
C2	Fragmented profile	20.139	10.033	20	1.700	1.063	19
C3	Indecisive profile	29.722	8.438	53	2.000	0.860	43
C4	Scientific profile	33.264	8.273	36	2.300	0.686	46

Note. IMTS scores and grades are coded so that higher scores reflect higher competence.

Persons differing in their knowledge profiles at T4 also differed systematically in their grades on the human memory course, $H(3) = 16.122$, $p = 0.001$. As with the IMTS scores, Jonckheere's test was significant, even though the median grades of students with the misconceptions profile C1 were slightly better than those of students with the fragmented profile C2. The other medians were in the assumed developmental rank order, which indicated that there was a significant trend from the fragmented profile, over the indecisive profile, to the scientific profile, $J = 2714.500$, $z = 3.547$, $r = 0.335$, which supports our hypothesis H5.

4. Discussion

4.1. Knowledge profiles

Our aim was to investigate whether conceptual change is still a relevant learning process in higher education students, therefore we conducted the first latent profile transition analysis in this population. In line with our Hypothesis 1, we found that there were individual differences in Psychology students' knowledge about human memory. Our analyses show that these individual differences of the 137 learners can be described in terms of only four knowledge profiles.

The misconceptions profile, in which students agreed mostly with static storage statements, was found in < 10% of the sample at all four measurement points, respectively. This does not mean that the students in the sample started the study with perfect prior knowledge, though. The most frequent profile at the beginning of the first semester was the fragmented profile characterized by strong agreement with both static storage statements and incorrect processing statements. These two kinds of statements are mutually incompatible, because either knowledge is static or it is processed in memory. We therefore interpreted students' agreement with both kinds of statements as indicating fragmented knowledge, in which the students hold both kinds of beliefs but do not understand their inter-relation. In line with Hypotheses 3a and 3b, the fragmented profile occurred frequently at the beginning of the study, decreased in its frequency over time, but was still found for 17% of the sample at the end of the study.

Many other students seemed to be aware of their lack of a complete understanding of memory, as the indecisive profile was the second most frequent profile at the first and the last measurement point, and the most frequent profile at the second and third measurement point. Students with this profile tended to choose the middle category when evaluating statements, neither agreeing nor disagreeing with them. Thus, not all students in our sample held deeply entrenched misconceptions. Instead, some students were aware of their lack of understanding. Students with the scientific profile agreed mostly with correct processing statements. They did not only understand that memory actively constructs and re-constructs information but also in which situations chunking, interference, or source monitoring are the most relevant memory processes.

The strongly decreasing sample proportions (from 56% at T1 to 17% at T4) for the fragmented profile and the strongly increasing sample proportions for the scientific profile (from 0% to 42%) demonstrated a trend towards more scientifically correct and also more integrated knowledge over the course of the four semesters (Hypotheses 2b and 3b). These findings with respect to the knowledge profiles and their changing frequencies over time are generally well in line with the findings obtained by M. Schneider and Hardy (2013) as well as Edelsbrunner, Schalk, Schumacher, and Stern (2015) with elementary school children. These two studies likewise found learners with fragmented knowledge and learners with integrated knowledge in their sample. Also, they found that fragmented knowledge decreased in its frequency in the sample but remained in some learners, and found an overall trend from misconceptions and fragmented knowledge to more integrated and correct knowledge. The main difference between the previous and the present findings is that the most frequent profile before learning was the misconceptions profile in the elementary school children and the fragmented profile in the university students. Future studies will have to investigate whether this is a general difference between the two age groups. Possibly, the longer learning history of the university students led to a greater amount of accumulated fragmented knowledge in comparison to the school children.

4.2. Transition paths

Students' knowledge changes over the four measurement points could be described in terms of few transition paths between the four knowledge profiles. There were only six paths taken by at least 5% of the sample, respectively, and a total of 81% of the sample was on these paths. The remaining part of the sample was on one of eight additional paths. The participants on the six most common paths either stayed in their respective latent profile or moved to a higher ranking profile, as shown in Fig. 3. Thus, there was a clear developmental ordering of the profiles from the misconceptions profile, over the fragmented profile and the indecisive profile, to the scientific profile. The knowledge profiles differed in how strongly participants agreed with statements about static storage concepts, incorrect processing concepts, or correct processing concepts of human memory. We thus interpret transitions between these knowledge profiles as evidence of conceptual change. The involved profiles demonstrate the strength of conceptual change underlying the transition: transitions between more similar profiles (e.g., the misconceptions profile and the indecisive profile) arguably require less knowledge restructuring than transitions between dissimilar profiles (e.g., the misconceptions profile and the scientific profile).

The highly systematic pathways imply that the learners' knowledge at each point in time is a good predictor of the learners' knowledge at later points in time. Indeed, we found significant and very strong relations between the knowledge profiles at different points in time. The fact that on the six most common paths more than half of the participants transitioned from one profile to another, that is, restructured their knowledge, shows that conceptual change is a relevant learning mechanism in higher education and in the domain of Psychology. Zero percent of the participants had the scientific profile at the start of the study. The 42% of the participants having this profile at the end of the study all transitioned there from lower-ranking knowledge profiles over time. This demonstrates that conceptual change does not only happen incidentally but is a central learning mechanism in acquiring academic concepts in higher education.

The findings also demonstrate the importance of knowledge fragmentation and integration for understanding and predicting conceptual change in higher education. As predicted, knowledge fragmentation was frequent at the start of the study, stayed constant on one path (the enduring fragmentation path), and decreased in its frequency on other paths (e.g., the decreasing fragmentation path and the slowly evolving scientific concepts path).

The findings demonstrate systematic associations of the profiles and pathways with domain-specific instruction, a finding also observed in the study by M. Schneider and Hardy (2013). Overall, 58% of the persons on the most frequent six paths transitioned to a different profile and 33% integrated previously fragmented knowledge during the first semester, in which most participants also attended a lecture about human memory. In contrast, only 11% of the participants transitioned to a different profile during the second, third, and fourth semester together, and no participants integrated their knowledge during that time.

4.3. Validity of the model results and limitations

At least five findings indicated that the model results reflect systematic and meaningful individual differences rather than random measurement error. First, the transition paths and the predictive relations between the knowledge profiles show a high degree of systematic organization in the results, both for inter-individual differences at each point in time and intra-individual differences over time. Second, the basic pattern of results in the current study conceptually replicates the key findings of M. Schneider and Hardy (2013), such as the small numbers of profiles and pathways, the co-existence of fragmented and integrated knowledge in the sample, and general trends towards more integrated and correct knowledge. Third, transitions to higher-ranking profiles were more frequent during domain-specific instruction (i.e., the lecture on human memory during the first semester) than at other times. Fourth, the latent profiles differed in their median rank-ordered grades, with students in higher-ranking profiles also showing better grades. The hypothesized rank order was not perfectly repoduced by the profiles, as participants with a misconceptions profile showed better grades than participants with a fragmented profile. However, this could be due to the fact that there were only seven individuals with a misconceptions profile and does not mean that a misconceptions profile per se is associated with better performance. Though evidence from analyses with most likely profile variables can be somewhat limited when entropy is not perfect, we believe that in our study, in which entropy was only slightly smaller than 0.90, such bias was only a marginal problem. Therefore, the findings support the criterion validity of the latent profile variables and demonstrate the grade-relevance of academic concepts and conceptual change in higher education. Finally, students in higher-ranking profiles also showed better performance on the IMTS, indicating that a better understanding of memory as involving the construction and re-construction of information went along with a greater skepticism with respect to the contents of autobiographical memories.

We expect that the findings of our study are highly generalizable, as the interpretation of our knowledge profiles is in line with previous studies conducted outside higher education (Edelsbrunner et al., 2015; McMullen et al., 2015; M. Schneider & Hardy, 2013). Our findings are somewhat limited as the identification of knowledge profiles in latent transition analysis is exploratory. There are no definite standards in choosing the number of profiles. In latent transition analysis, the number of estimated parameter rises exponentially with the number of profiles assumed. With 63 free parameters in our model with four profiles at four measurement points and 137 participants, our design is at the margin of being underpowered. Our sample is restricted to psychology students that represent a highly selected, presumably high-achieving subpopulation. Future studies should investigate whether students from other programs than psychology show similar knowledge profiles and developmental patterns. Another limitation is the use of a newly constructed test in our study. Future studies will have to further examine the validity and reliability of the test and its sensitivity to knowledge changes over time.

4.4. Theoretical, methodological, and practical implications

Our findings have theoretical, methodological, and practical implications. On the theoretical level, the results demonstrate that conceptual change still takes place in higher education students and is vital for the acquisition of an understanding of academic concepts. Conceptual change, its measurement, its prevalence, its dynamics over time, and its stimulation by learning environments are well investigated for school students but almost not examined at all for students in higher education. Our present findings suggest that future studies should systematically investigate parallels and possible differences between conceptual change in K-12 school and in higher education. This will lead to a better understanding of which processes are specific to conceptual change in general and which processes only occur in specific age groups. Further, future studies should include motivational and affective covariates into latent transition models in order to investigate non-cognitive determinants of conceptual change. The small number of profiles and pathways indicated that individual differences matter in conceptual change, but that the number of possible individual differences is quite limited. This suggests that the learners' idiosyncratic knowledge construction processes are constrained by cognitive and environmental variables which guide these processes along a few developmental trajectories.

On the methodological level, our findings indicate that latent profile transition analyses are useful tools for modeling longitudinal changes in multidimensional knowledge structures in general, and specifically for modeling conceptual change. Only very few latent profile transition analyses have been published so far (see Hickendorff et al., this issue). The convergence of findings from earlier studies with school children and the current study with Psychology students indicates a high degree of replicability, stability, and generalizability of the findings across studies, age groups and content domains. Latent profile transition analyses have also proven to be effective data reduction techniques, because they help to describe individual differences in large samples in terms of just a few profiles and transition paths.

On the practical level, findings from studies like ours can inform higher education teaching. Teachers in higher education might often lack awareness that students enter their programs with prior beliefs and misconceptions, because higher education learning has long been neglected in research on conceptual change. The profiles and pathways identified in the current study can help teachers to better identify students' current knowledge and to predict students' future pathways of learning. Instruction can thus be tailored to fit subpopulations of students differing in their prior knowledge or development. The current study as well as the study by M. Schneider and Hardy (2013) also found evidence of systematic relations between instruction and the knowledge profiles and pathways. Future studies should investigate these interactions in greater detail.

Appendix A

Latent transition probabilities based on the estimated model: T1 classes (rows) by T2 classes (columns)

Class		C1	C2	C3	C4
C1	Misconceptions profile	0.709	0.000	0.070	0.221
C2	Fragmented profile	0.000	0.389	0.611	0.000
C3	Indecisive profile	0.000	0.000	0.319	0.681
C4	Scientific profile	0.000	0.000	0.000	1.00

Latent transition probabilities based on the estimated model: T2 classes (rows) by T3 classes (columns)

Class		C1	C2	C3	C4
C1	Misconceptions profile	1.00	0.000	0.000	0.000
C2	Fragmented profile	0.049	0.786	0.165	0.000
C3	Indecisive profile	0.000	0.000	0.856	0.144
C4	Scientific profile	0.000	0.000	0.000	1.00

Latent transition probabilities based on the estimated model: T3 classes (rows) by T4 classes (columns)

Class		C1	C2	C3	C4
C1	Misconceptions profile	0.630	0.000	0.037	0.000

C2	Fragmented profile	0.000	0.867	0.133	0.000
C3	Indecisive profile	0.000	0.039	0.765	0.196
C4	Scientific profile	0.000	0.000	0.000	1.00

Appendix B

Example task "chunking"

Wild horses

Two groups of children participate in the study.

The children in one of these groups (Group A) are twelve year-olds, who know little about horses in general.

The children in the other group (Group B) are eight year-olds, who know a lot about horses in general.

The two groups do not differ in terms of the children's intelligence and the number of boys and girls in the group.

All participants read a simple text about wild horses that teaches them what kinds of species exist, where they are found and what they need to live. Afterwards, the children are asked to answer questions assessing their knowledge and comprehension, based on the text from memory.

How sure are you that each of the following statements is correct or incorrect?

	Definitely incorrect Definitely correct
Group A will perform markedly better than Group B because they have four more years of practice in reading and remembering from texts than the younger children.	o o o o o o o
Group B will perform markedly worse than Group A, because the children are on a lower stage of cognitive development and therefore cannot process information so well.	o o o o o o o
Group B will perform markedly better than Group A because they have more prior knowledge and therefore can store information from the text in a more structured way.	o o o o o o o
Group A will perform markedly worse than Group B because the memory of twelve year-olds is partly impaired already due to hormonal changes during puberty.	o o o o o o o
Both groups will perform almost equally well because they have read the same text with the result that each person has memorized the same information.	o o o o o o o
There barely will be any difference between the two groups because their intelligence is similar and therefore they can process information almost equally well.	o o o o o o o

Example task "interference"

Font sizes

In one study the participants are presented with 50 pairs of simple numbers successively on a computer screen, for examples 4 and 9. One number is always presented in a smaller font size and the other one in a larger font size. The participants' task for each pair of numbers is to answer as quickly as possible whether the left or the right number is presented in a larger font size.

For half of the participants (Group A) the number with the larger numeric value (e.g. 9) is always presented in a larger font size and the number with the smaller numeric value (e.g. 4) in a smaller font size.

For the other half of the participants (Group B) the number with the smaller numeric value (e.g. 4) is always presented in a larger font size and the number with the larger numeric value (e.g. 9) in a smaller font size.

It is analyzed how many milliseconds participants in Group A and Group B need on average to evaluate which one of the two numbers is presented in a larger font size.

How sure are you that each of the following statements that is correct or incorrect?

	Definitely incorrect Definitely correct
Group A will be markedly faster than Group B because the numeric values and font sizes in Group B provide the brain with contradicting information.	o o o o o o o
Group B will be markedly slower than Group A because people are used to larger numeric values being presented in larger font sizes than smaller numeric values in everyday life.	o o o o o o o
	o o o o o o o

Group B will be markedly faster than Group A because it is more distinct when numeric values and font sizes do not match and therefore it is remembered more easily.	
Group A will be markedly slower than Group B because the task of Group B is more unusual and therefore more interesting and motivating.	o o o o o o o
Both groups will perform almost equally fast because the brain processes numeric values and font sizes independently.	o o o o o o o
There barely will be any difference between the two groups because the task is about font sizes only and the participants will pay no attention to the numbers.	o o o o o o o

Example task "source monitoring"

Europe

Participants in a study know little about the European economic system. They read a text on the topic "Should all Europeans be allowed to live and work anywhere in the European Union without restrictions?" Participants are told that the author of the text is a participant in a casting show for singing talents and has little expertise in politics. The arguments in the text are partly true and partly incorrect.

The participants rate, right after reading the text, how convincing the arguments are (Test 1). One month later, they are presented with the arguments a second time without announcement and again, they are asked again how convincing they find the arguments (Test 2).

How sure are you that each of the following statements is correct or incorrect?

	Definitely incorrect ... Definitely correct
The arguments will be judged to be clearly more convincing in Test 1 than in Test 2 because they will be more available in memory during Test 1.	o o o o o o o
The arguments will be judged to be clearly less convincing in Test 2 than in Test 1 because the participants will be less impressed by the author's fame.	o o o o o o o
The arguments will be judged to be clearly more convincing in Test 2 because participants will have spent more time thinking about the topic and will assume a more positive attitude.	o o o o o o o
The arguments in Test 1 will be judged to be clearly less convincing than in Test 2 because participants will be more likely to remember in Test 1 that the arguments were from an author with little expertise in politics.	o o o o o o o
The arguments are judged in a comparable way in both tests because the arguments did not change in the meantime.	o o o o o o o
The judgment of the arguments will barely differ between the two tests because attitudes towards political topics are embedded deeply in memory.	o o o o o o o

References

American Psychological Association (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*(12), 1060–1073. http://dx.doi.org/10.1037/0003-066X.57.12.1060.

Asparouhov, T., & Muthén, B. (2010). *Computing the strictly positive Satorra-Bentler chi-square test in Mplus. Mplus Web Notes, Vol. 12*.

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2009). *Memory.* Hove: Psychology Press.

Beeth, M., Duit, R., Prenzel, M., Ostermeier, C., Tytler, R., & Wickman, P.-O. (2003). Quality development projects in science education. In D. Psillos, P. Kariotoglou, V. Tselfes, E. Hatzikraniotis, G. Fassoulopoulos, & M. Kallery (Eds.). *Science education research in the knowledge-based society* (pp. 447–457). Dordrecht: Springer Netherlands.

Bjork, R. A. (1992). Interference and memory. In L. R. Squire (Ed.). *Encyclopedia of learning and memory* (pp. 283–288). New York: Macmillan.

Brooks, L. R. (1967). The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology, 19*(4), 289–299. http://dx.doi.org/10.1080/14640746708400105.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Carey, S. (1992). The origin and evolution of everyday concepts. In R. Giere (Ed.). *Cognitive models of science* (pp. 89–128). Minneapolis: University of Minnesota Press.

Chase, W. G., & Simon, H. A. (1973). Percpetion in chess. *Cognitive Psychology, 4*, 55–81.

Clark, S. L., & Muthén, B. (2009). *Relating latent class analysis results to variables not included in the analysis.* Unpublished paper available at www.statmodel.com/download/relatinglca.pdf.

diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science, 28*, 843–900.

Duit, R., Treagust, D. F., & Widodo, A. (2008). Teaching science for conceptual change: Theory and practice. In S. Vosniadou (Ed.). *International handbook of research on conceptual change* (pp. 629–646). New York: Routledge.

Edelsbrunner, P. A., Schalk, L., Schumacher, R., & Stern, E. (2015). Pathways of conceptual change: Investigating the influence of experimentation skills on conceptual knowledge development in early science education. *Paper presented at the Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 620). .

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208*, 1181–1182.

European Commission (2008). *The European Qualifications Framework for Lifelong Learning (EQF).* Luxembourg: Office for Official Publications of the European Communities.

Garry, M., Loftus, E. F., & Brown, S. W. (1994). Memory: A river runs through it. *Consciousness and Cognition, 3*(3), 438–451.

Gentner, D., Brem, S., Ferguson, R. W., Markman, A. B., Levidow, B. B., Wolff, P., & Forbus, K. D. (1997). Analogical reasoning and conceptual change: A case study of Johannes Kepler. *Journal of the Learning Sciences, 6*(1), 3–40.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*, 236–243.

Goff, L. M., & Roediger, H. L. (1998). Imagination inflation for action events: Repeated imaginings lead to illusory recollections. *Memory & Cognition, 26*(1), 20–33. http://dx.doi.org/10.3758/BF03211367.

Goldstone, R. L., & Kersten, A. (2003). Concepts and categorization. In I. B. Weiner (Vol. Ed.), *Handbook of psychology: Vol. 4*, (pp. 599–621). Hoboken, NJ: Wiley.

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking". *Journal of Educational Psychology, 98*(2), 307–326.

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relations between physical and semantic size in comparison tasks. *Memory & Cognition, 10*(4), 389–395.

Hewson, P. W., Tabachnick, B. R., Zeichner, K. M., & Lemberger, J. (1999). Educating prospective teachers of biology: Findings, limitations, and recommendations. *Science Education, 83*(3), 373–384. http://dx.doi.org/10.1002/(SICI)1098-237X(199905)83:3 < 373::AID-SCE6 > 3.0.CO;2-3.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly, 15*(4), 635–650. http://dx.doi.org/10.1086/266350.

Johnson, M. K., Hashtroudi, S., & Lindsay, S. D. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3–28.

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika, 41*(1/2), 133–145. http://dx.doi.org/10.2307/2333011.

Kainulainen, M., McMullen, J., & Lehtinen, E. (2017). Early developmental trajectories toward concepts of rational numbers. *Cognition and Instruction, 35*. http://dx.doi.org/10.1080/07370008.2016.1251287.

Linn, M. C. (2006). The knowledge integration perspective on learning and instruction. In R. K. Sawyer (Ed.). *The Cambridge handbook of the learning sciences* (pp. 243–264). New York: Cambridge University Press.

Lix, L. M., & Keselman, H. J. (1998). To trim or not to trim: Tests of location equality under heteroscedasticity and nonnormality. *Educational and Psychological Measurement, 58*(3), 409–429.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7*(4), 560–572. http://dx.doi.org/10.1016/0010-0285(75)90023-7.

Lynn, S. J., & McConkey, K. M. (Eds.). (1998). *Truth in memory*. New York: Guilford.

Machery, E. (2010). Précis of Doing without Concepts. *Behavioral and Brain Sciences, 33*, 195–244.

McMullen, J., Laakkonen, E., Hannula-Sormunen, M., & Lehtinen, E. (2015). Modeling the developmental trajectories of rational number concept (s). *Learning and Instruction, 37*(1), 14–20. http://dx.doi.org/10.1016/j.learninstruc.2013.12.004.

Melcher, J. M., & Schooler, J. W. (1996). The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory and Language, 35*(2), 231–245. http://dx.doi.org/10.1006/jmla.1996.0013.

Merz, C. J., Dietsch, F., & Schneider, M. (2016). The impact of psychosocial stress on conceptual knowledge retrieval. *Neurobiology of Learning and Memory, 134*, 392–399. http://dx.doi.org/10.1016/j.nlm.2016.08.020.

Muthén, L. K., & Muthén, B. O. (1998–2016). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Niedźwieńska, A., Neckar, J., & Baran, B. (2007). Development and validation of the implicit memory theory scale. *European Journal of Psychological Assessment, 23*(3), 185–192. http://dx.doi.org/10.1027/1015-5759.23.3.185.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535–569. http://dx.doi.org/10.1080/10705510701575396.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211–227. http://dx.doi.org/10.1002/sce.3730660207.

Potvin, P., Masson, S., Lafortune, S., & Cyr, G. (2015). Persistence of the intuitive conception that heavier objects sink more: A reaction time study with different levels of interference. *International Journal of Science and Mathematics Education, 13*(1), 21–43. http://dx.doi.org/10.1007/s10763-014-9520-6.

Rittle-Johnson, B., Schneider, M., & Star, J. (2015). Not a one-way street: Bi-directional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*, 587–597. http://dx.doi.org/10.1007/s10648-015-9302-x.

Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impact of undergraduate science course innovations on learning. *Science, 331*(6022), 1269–1270. http://dx.doi.org/10.1126/science.1198976.

Satorra, A., & Bentler, P. M. (2010). Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika, 75*(2), 243–248. http://dx.doi.org/10.1007/s11336-009-9135-y.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? *Methodology, 6*(4), 147–151. http://dx.doi.org/10.1027/1614-2241/a000016.

Schneider, M. (2012). Knowledge integration. In N. M. Seel (Ed.). *Encyclopedia of the sciences of learning* (pp. 1684–1686). New York: Springer.

Schneider, M., & Hardy, I. (2013). Profiles of inconsistent knowledge in children's pathways of conceptual change. *Developmental Psychology, 49*(9), 1639–1649. http://dx.doi.org/10.1037/a0030976.

Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology, 81*(3), 306–312.

Schneider, M., Vamvakoussi, X., & Van Dooren, W. (2012). Conceptual change. In N. M. Seel (Ed.). *Encyclopedia of the sciences of learning*. New York: Springer.

Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science, 8*, 118–137. http://dx.doi.org/10.1111/tops.12174.

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition, 124*, 209–215.

Terpstra, T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae, 14*(3), 327–333.

Tynjälä, P. (1999). Towards expert knowledge? A comparison between a constructivist and a traditional learning environment in the university. *International Journal of Educational Research, 31*(5), 357–442. http://dx.doi.org/10.1016/S0883-0355(99)00012-9.

Vosniadou, S. (2008). *International handbook for research on conceptual change.* New York: Routledge.

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., ... Blakemore, S.-J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition, 22*, 264–271. http://dx.doi.org/10.1016/j.concog.2013.01.004.