# Assessing prior knowledge types as predictors of academic achievement in the introductory phase of biology and physics study programmes using logistic regression

Torsten Binder[1*] , Angela Sandmann[1], Bernd Sures[2], Gunnar Friege[3], Heike Theyssen[4] and Philipp Schmiemann[1]

## Abstract

**Background:** Increasingly, high dropout rates in science courses at colleges and universities have led to discussions of causes and potential support measures of students. Students' prior knowledge is repeatedly mentioned as the best predictor of academic achievement. Theory describes four hierarchically ordered types of prior knowledge, from declarative *knowledge of facts* to procedural *application of knowledge*. This study explores the relevance of these four prior knowledge types to academic achievement in the introductory phase of the two science subjects, biology and physics.

**Results:** We assessed the knowledge types at the beginning and student achievement (measured by course completion) at the end of the first study year. We applied logistic regression models to evaluate the relationship between the knowledge types and academic achievement. First, we controlled for a well-established predictor of academic achievement (high school grade point average). Second, we added the knowledge types as predictors. For biology, we found that only knowledge about principles and concepts was a significant predictor in the first year. For physics, knowledge about concepts and principles as well as the ability to apply knowledge to problems was related to academic achievement.

**Conclusion:** Our results concerning the knowledge types, which are of special relevance in biology and physics studies, could lead to effective measures, e.g. for identifying at-risk students and course guidance. Furthermore, the results provide a profound starting point for controlled intervention studies that systematically foster the identified relevant knowledge types in each subject and aim at a theory- and empirical-based optimization of pre- and introductory courses.

**Keywords:** Biology, Physics, Higher education, Academic achievement, Knowledge types

## Introduction

Higher education research has always been concerned with the prediction of academic achievement to improve teaching and enhance the self-awareness of students (Kappe & van der Flier, 2012; Schiefele, Krapp, & Winteler, 2014; Thompson & Zamboanga, 2003). In recent years, the prediction of academic achievement in science-related fields in particular has gained increasing interest due to high dropout rates and the declining numbers of students majoring in these fields (Chen & Soldner, 2013; Eurostat, 2016; Olson & Riordan, 2012; Whalen & Shelley, 2010). In particular, the first year at university shows a substantial number of science-leavers (Alting & Walser, 2007; Chang, Cerna, Han, & Saenz, 2008). The identification of good predictors of first-year grades could lead to effective support of at-risk students by adjusted training and course guidance. This applies even more, since first-year grades themselves have proven to be good predictors of the final grades (Adelman, 1999; Harackiewicz, Barron, Tauer, & Elliot, 2002). A variety of underlying factors are discussed

as reasons for science dropout or achievement, such as *demographics*, *personal factors*, *attitudinal factors*, and *cognitive factors* (for an overview see, Chen & Soldner, 2013; Robbins et al., 2004). One of the most powerful *cognitive* predictors of academic achievement in general is prior knowledge (Hell, Trapmann, & Schuler, 2007; Ramist, Lewis, & McCamley-Jenkins, 2001).

In this regard, Dochy, Segers, and Buehl (1999) highlighted the role of a more detailed and complete prior knowledge assessment with a variety of assessment methods. Thus, we have developed and tested prior knowledge assessments in biology and physics that differentiate four prior knowledge types in each subject. We investigated in how far the results of these assessments predict academic achievement in the first year of biology and physics majors.

## Theoretical background

### Predicting academic achievement in science subjects

In recent years, a variety of predictors of success and retention in science courses were identified. In single studies, predictors such as work-study aid, participation in learning communities, motivation, personality traits, college admission tests, and subject-specific prior knowledge were identified (e.g. prior knowledge: Hailikari, Nevgi, & Lindblom-Ylänne, 2007; van Riesen, Gijlers, Anjewierden, & Jong, 2018; concept inventories and attitude: Lee, Sbeglia, Ha, Finch, & Nehm, 2015; SAT and ACT: Sadler & Tai, 2007; personal, institutional, cognitive and affective predicators: Whalen & Shelley, 2010; Sadler & Tai, 2001).

However, of these identified predictors, the most powerful seem to be the cognitive ones, prior knowledge in particular. Meta-analyses (Hell et al., 2007; Kuncel, Hazlett, & Ones, 2001; Robbins et al., 2004) identified prior knowledge as the best predictor of academic success. Most studies used the high school grade point average (HS GPA) as an indicator of students' prior knowledge. Shown by various meta-analyses (e.g. Burton & Ramist, 2001; Hell et al., 2007; Ramist et al., 2001; Robbins et al., 2004), the HS GPA is one of the best universal predictors of achievement in university courses. Not only it is a measure for prior knowledge but also for university readiness, and its predictive power for academic achievement exceeds that of college admission tests (e.g. ACT), when considered together (Chingos, 2018). However, the HS GPA is rather subject-unspecific. Almost the same applies for the high school course choice in mathematics or mathematics tests scores (e.g. SAT-M) that have proven to be relevant predictors for academic success in science subjects (Burton & Ramist, 2001, e.g. biology: Loehr, Almarode, Tai, & Sadler, 2012; Sadler & Tai, 2007; physics: Hazari, Tai, & Sadler, 2007; Sadler & Tai, 2007). In order to identify more subject-specific predictors for academic success

in science subjects, several other, more or less explicit, indicators for prior knowledge were used (Table 1).

To predict achievement in biology courses, *high school biology enrolment* or *biology high school grades* were used as indicators for subject-specific prior knowledge in some studies. Some studies use concept inventories (Table 1) that are closely related to certain content areas. Therefore, in these studies, the test scores are only correlated with the respective courses' grades. One large-scale, multi-level modelling study (Loehr et al., 2012) addressed, among others, *high school science grades*, *laboratory pedagogy*, and *how biology content was covered in high school* as predictors of the final grade in introductory biology. They found that students who took a high school course with a focus on a deep understanding of biology content performed a third of a grade better in their first introductory biology course than their peers who took a course focusing on memorising facts. Thus, in predicting biology achievement at university, it seems promising to address in which way biology content was packaged and learned by the students at high school.

Most physics prediction studies used implicit indicators for physics prior knowledge. For example, Sadler and Tai (2001) employed a comprehensive set of demographic, student, and teacher variables to predict physics achievement for a sample of $N = 1,933$ introductory college physics students. In their study, those variables that are implicitly connected to physics prior knowledge turned out to be among the most powerful predictors for the introductory college grade (e.g. *regular*, *honours*, or *AP physics*, *years of physics in high school*). Some studies use physics tests as an indicator for students' prior knowledge. One of the first studies using a physics test was a study by Halloun and Hestenes (1985). They designed a Mechanics Diagnostic Test (MDT) which assessed the qualitative understanding of basic physics concepts. The test score of the MDT was the best predictor for the exam score in introductory physics, even when controlling for other indicators of prior knowledge (*mathematics test*, *physics courses*, and *mathematics courses in high school*) (Halloun & Hestenes, 1985). Sorge, Petersen, and Neumann (2016) found similar results. In their study a physics prior knowledge test outperformed the HS GPA and the high school physics grade as a predictor for physics achievement at university.

In conclusion, the HS GPA is a useful predictor of academic achievement in different subjects (cf. Robbins et al., 2004). However, it is rather subject-unspecific and combines knowledge of subjects with general knowledge and cognitive abilities. This suggests that prediction studies should consider the HS GPA as a control variable in order to find more precise and specific predictors for achievement.

Binder *et al. International Journal of STEM Education*        (2019) 6:33

Page 3 of 14

**Table 1** Overview of significant cognitive predictors for academic achievement in science courses at university or college

| Subject | Biology or physics in high school | Subject-specific grade in high school | Number of lab courses | Subject-specific tests in biology or physics | Reference |
|---------|-----------------------------------|---------------------------------------|----------------------|----------------------------------------------|-----------|
| Biology | ✓ (Number of AP courses, enrolment in AP biology) | ✓ (Science) | ✓ | – | Loehr et al. (2012) |
| | – | ✓ (Science, AP exam score) | – | – | Sadler & Tai (2007) |
| | – | – | – | ✓ (CINS, ACORNS) | Lee et al. (2015) |
| | – | – | – | ✓ (GCA) | Smith, Wood, & Knight (2008) |
| Physics | ✓ ‡ (Physics enrolment) | – | – | ✓ (Mechanics test) | Halloun & Hestenes (1985) |
| | ✓ (Physics enrolment) | – | – | – | Hart & Cottle (1993) |
| | ✓ (Physics enrolment) | – | – | – | Alters (1995) |
| | ✓ (At least 2 years of physics, enrolment of any physics course) | ✓ (Physics) | ✓ | – | Sadler & Tai (2001) |
| | – | ✓ (Science, AP exam score) | – | – | Sadler & Tai (2007) |
| | – | – | – | ✓ (FCI) | Docktor & Heller (2008) |
| | – | ✗ (Physics) | – | ✓ (Physics test) | Sorge, Petersen, & Neumann (2016) |
| | – | – | – | ✓ (Physics tests) | Buschhüter, Spoden, & Borowski (2017) |
| | ✓ (Time on mechanics and optics) | ✓ (Science) | ✓ (Microlabs) | – | Hazari, Tai, & Sadler (2007) |

*CINS* Concept Inventory on Natural Selection (Anderson, Fischer, & Norman, 2002), *ACORNS* Assessment of Contextual Reasoning on Natural Selection (Nehm, Beggrow, Opfer, & Ha, 2012), *GCA* Genetics Concept Assessment (Smith, Wood & Knight, 2008), *FCI* Force Concept Inventory (Hestenes, Wells, & Swaghammer, 1992)
‡Variable only significant for the college sample, but not for the university sample
✗Variable not significant
✓Variable significant

In addition to HS GPA, a variety of indicators for subject-specific prior knowledge have been used for the prognosis of achievement in science. In both science subjects (biology and physics), indictors of subject-specific prior knowledge, such as high school science grades or *course enrolment*, seem to be related to achievement. Those indicators only address prior knowledge indirectly and do not measure the knowledge relevant for the courses in the first year.

A few studies address prior knowledge directly, using prior knowledge tests. In these studies, prior knowledge is one of the best predictors for future physics achievement. However, the findings are only based on a few studies and more evidence is needed.

Nevertheless, some results in biology and physics (e.g. Loehr et al., 2012; Halloun & Hestenes, 1985) present indications that certain types of prior knowledge (e.g. deep understanding) could be better predictors for academic achievement than other types (e.g. memorised facts). Therefore, a systematic comparison of different prior knowledge types, based on a theoretical model of prior knowledge, as predictors for achievement seems beneficial to us. Studies describing the relationship between prior knowledge types and academic achievement will be discussed in the next section.

**Knowledge types as predictors of academic achievement**
The relevance of prior knowledge to learning is familiar to psychologists and educators. In particular, domain-specific prior knowledge is argued to be crucial for the acquisition of new knowledge (Bloom, 1976; Dochy, 1992, 1994; Krathwohl, 2002). High and well-developed prior knowledge about a topic supports learning and vice-versa (Ausubel, 2000; Schneider & Pressley, 1997; Thompson & Zamboanga, 2003). Thus, individuals with higher prior knowledge about a topic understand and

remember the subject matter better and perform better on exams (Chi & Ceci, 1987; Hailikari et al., 2007).

Most studies use one ability measure as an indicator of the overall prior knowledge of a student. In contrast to these studies, Hailikari et al. (2007) and Bloom (1976) define prior knowledge as composed of different types of knowledge, skills, and competencies, multidimensional and dynamic in nature. Following general theories about knowledge, they assume that not all prior knowledge types influence academic achievement the same way.

Hailikari et al. (2007) introduce a structural model of prior knowledge that distinguishes between four prior knowledge types (*knowledge of facts*, *knowledge of meaning*, *integration of knowledge*, and *application of knowledge*; Fig. 1). The model is based on the theories of knowledge and learning described in the revision of educational objectives by Krathwohl (2002), studies of knowledge dimensions by Dochy (1992), and the theory of structurally developing understanding by Biggs (1996).

Knowledge of facts is defined as knowledge on a low level of abstraction, which can be tested with simple recognition and reproduction tasks. Knowledge of meaning is the ability to understand the meaning of a concept. Integration of knowledge is the ability to understand the links and interrelations of concepts and different phenomena. For application of knowledge, students must be able to apply knowledge and solve domain-specific problems (for details of the definitions, see Hailikari et al., 2007).

Every type of knowledge is ordered into the widely accepted distinction between declarative and procedural knowledge (Anderson, 1982; Birenbaum & Dochy, 2012). Declarative knowledge types (knowledge of facts and knowledge of meaning) are closely related to cognitive indicators for retention (remember, recall, recognise, reproduce, etc.). The procedural types (integration of knowledge and application of knowledge) are closely related to cognitive indicators for transfer (understand, apply, etc.) (cf. Mayer, 2002). These indicators were adopted from Bloom's original Taxonomy of Learning (Bloom, 1976).

The model describes the structure of knowledge and not primarily the development of knowledge or learning. It is meant to serve as a framework for assessment and for the prediction of academic achievement (Hailikari et al., 2007). Therefore, it could serve as a design objective for the construction of ample knowledge tests. But with regard to the analysis of learning and interventions that might be derived from the results of prediction studies, it is important that the model is compatible with recent theories of knowledge development, e.g. of knowledge integration (e.g. Geller, Neumann, Boone, & Fischer, 2014; Liu, Ryoo, Linn, Sato, & Svihla, 2015). Hailikari et al. (2007) explicitly point out that in "the model it is assumed that operating on the higher levels of knowledge subsumes the lower levels of knowledge" (p. 324).

Based on this framework, Hailikari et al. (2007) investigated the relevance of the four prior knowledge types in
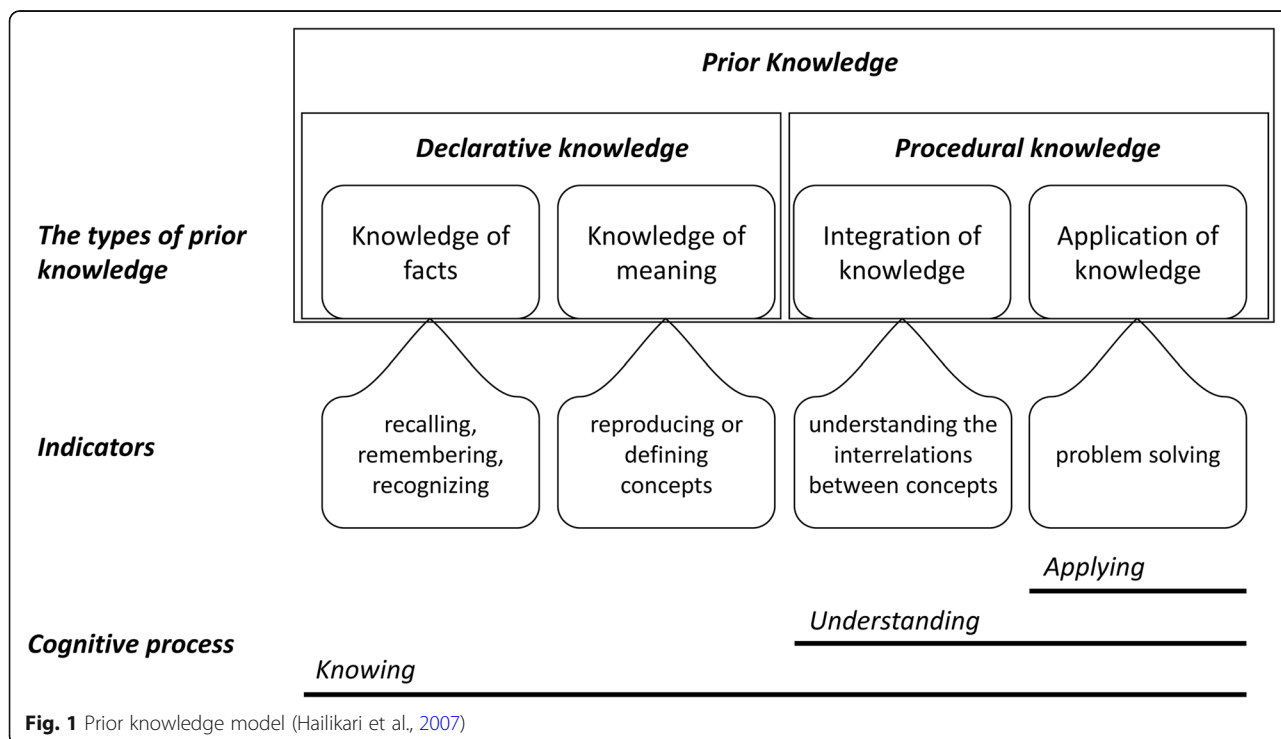


**Fig. 1** Prior knowledge model (Hailikari et al., 2007)

mathematics. Hailikari, Katajavuori, and Lindblom-Ylänne (2008) utilised the model to assess undergraduate pharmacology students, and Hailikari and Nevgi (2010) used it to identify at-risk chemistry students. In all three studies, procedural types of knowledge turned out to be the best predictors for course grades in the respective subjects. Thus, the model of prior knowledge devised by Hailikari et al. (2007) was successfully adapted to a variety of science fields and added a new viewpoint on prior knowledge as a predictor of academic achievement. However, in all studies, only a few items or one single item was used to assess a particular type of knowledge.

In addition, findings about the relevance of the prior knowledge types for academic achievement are hitherto hardly known. Until now, only correlations between the prior knowledge types and single exams in chemistry and mathematics were reported. Especially biology and physics courses impose diverse requirements on the students in their first year. In biology courses, much new taxonomy and terminology need to be learned in different content areas (e.g. botany), whereas in physics courses, the students are encouraged to solve problem-solving tasks in different content areas (e.g. mechanics). Therefore, different prior knowledge types could be relevant for achievement in the first year. Findings could be of immediate practical relevance to plan meaningful interventions and course-guidance for first-year students in both science subjects.

Besides the practical relevance, a transfer of the prior knowledge model by Hailikari et al. (2007) to biology and physics would legitimise the model for science domains besides chemistry and mathematics.

### Aims of the study
The goal of our study is to examine the relevance of different prior knowledge types for the prediction of academic achievement in the first year of biology and physics majors.

Concerning the role of the HS GPA as a well-established predictor across subjects, our study aims to replicate the results of former studies and, if it is a relevant predictor, to use the HS GPA as a control variable for further analyses.

With regard to our main goal, we want to construct a detailed, reliable, and valid assessment of students' prior knowledge types, using the framework established by Hailikari et al. (2007). Applying this assessment, we want to identify the knowledge types that are most relevant to academic achievement in the first year at university in biology and physics. Therefore, we pose the following research questions:

I. In how far does the HS GPA predict the likelihood of having academic achievement in the first year at university for biology and physics majors?

II. In how far are the *knowledge types* incrementally valid compared to the HS GPA in predicting the likelihood of having academic achievement in the first year for biology and physics majors?

One can assume that a valid and reliable assessment of the knowledge types should produce more precise predictors for academic achievement than HS GPA. The identification of prior knowledge types related to academic achievement can be a starting point for knowledge-based course guidance and the identification of at-risk students.

## Methods
In our study, we sought to utilise knowledge types in biology and physics as predictors of academic achievement in these subjects. Therefore, we assessed the knowledge types as well as HS GPA at the beginning of the first semester and academic achievement at the end of the second semester. The following section explains how the variables were assessed and used for the prediction analysis.

### Study and test design
The study took place in the first year at two universities in Germany. We tested the students' prior knowledge in the first two weeks of the first semester. We administered tests for all four prior knowledge types (see below). All biology tests were administered to the biology students and all physics tests to the physics students. In addition, we asked for HS GPA as self-report. At the end of the first and second semester, we obtained students' course grades (see below).

### Prior knowledge types
In both subjects, biology and physics, tests for all four knowledge types were constructed with regard to the curriculum and the relevant cognitive indicators and processes of every type of knowledge.

#### Test construction
To construct the assessment instrument, we used the model of Hailikari et al. (2007) as a framework. They recommend using different assessment methods to measure different knowledge types. This argument is consistent with Dochy et al. (1999), who suggested using alternative, authentic forms of assessments that initiate adequate cognitive processes. Considering this, we applied different test formats for the four knowledge tests.

#### Test content regarding the curriculum
We analysed university curricula and common subject-specific textbooks of biology and physics to uncover relevant content areas for the first year. Recurring content areas were included in the assessment. In biology, three content areas are relevant for the introductory

courses: botany, zoology, and cellular biology. In physics, we identified mechanics and electrodynamics as main content areas. Thus, we focused on these content areas for test construction. To assure content validity, we referred to the curricula and textbooks and involved lecturers or training staff to assess the relevance and correctness of the items.

### Knowledge of facts
Hailikari et al. (2007) define knowledge of facts to be knowledge on a low level of abstraction that can be tested with simple reproduction tasks. Thus, we choose a multiple-choice (single-select) format to assess knowledge of facts. We designed or adapted a total of 106 items in biology and 106 items in physics in the content areas mentioned above. Some of the physics items were adapted from Müller, Fischer, Borowski, and Lorke (2017) and some of the biology items from Schachtschneider (2016). All of the biology and physics items asked for facts, such as technical terms, nomenclature, taxonomical names, or formulas (cf. Additional file 1).

After expert revision of the items (cf. the "Test content regarding the curriculum" section), a portion of the items (39 in biology, 15 in physics) was chosen for the final knowledge of facts assessment that covered the relevant facts of the two semesters. All items had one attractor and three distractors. We coded the answers dichotomously.

### Knowledge of meaning
Similar to knowledge of facts, knowledge of meaning is also on a low level of abstraction but deals with another quality of knowledge (cf. de Jong & Ferguson-Hessler, 1996). It includes the ability to reproduce the definitions of concepts and laws of a subject. Hailikari et al. (2007) suggest to use open questions in order to assess this knowledge type. Following this, we constructed 15 open-ended tasks for biology and physics each. Each task asked for a short definition of a subject-specific concept or law (cf. Additional file 1). We used a coding scheme defining meaningful aspects of all concepts to rate students' answers. The answers were coded based on the number of meaningful aspects within the student's concept description. Therefore, we applied polytomous coding.

### Integration of knowledge
Understanding the interrelationships among concepts in a domain is often referred to as *conceptual knowledge* (de Jong & Ferguson-Hessler, 1996; Mitchell & Chi, 1984), a part of declarative knowledge. Other authors view it as its own type of knowledge, termed *structural knowledge* (Jonassen, Yacci, & Beissner, 2012) or *propositional knowledge* (Ruiz-Primo & Shavelson, 1996). Again, we follow the argument of Hailikari et al. (2007), who view this type

of knowledge as procedural because of its active nature and the reasoning skills it requires when the students understand interrelationships between different concepts of a domain. As this process is closely related to the cognitive structure as defined in Ausubel's Assimilation Theory (Ausubel, 1963, 1968), we considered concept maps an appropriate assessment tool. For more than 25 years, concept maps have been a practical and reliable assessment tool to reveal cognitive structures (e.g. Novak & Gowin, 1999; Ruiz-Primo & Shavelson, 1996).

We used a *construct-a-concept-map* task (cf. Yin, Vanides, Ruiz-Primo, Ayala, & Shavelson, 2005) for each subject with 12 pre-structured concepts but without linking lines and labels. In this type of assessment, the students are required to understand the interrelations among the concepts, draw linking lines, and write down linking phrases between the concepts (cf. Additional file 1). The concepts are equally spread over the domains of the first and second semester (cf. the "Test content regarding the curriculum" section). A coding scheme was adopted to code the propositions (i.e. the linking phrase with regard to the linked concepts). Every proposition was treated as one single item of the test and scored dichotomously (1, right; 0, wrong or missing).

### Application of knowledge
Application of knowledge is shown by the ability to apply knowledge, for example in solving subject-specific problems (Hailikari et al., 2007). This type of knowledge is revealed when students are confronted with novel problems that require them to understand the problem and adapt a learned solution procedure (Hailikari et al., 2007). Based on these definitions, we evaluated the tasks given in the courses and designed sets of 12 problems each for biology and physics. The solution of each problem can be reached by the application of subject-specific concepts taught in the classes, such as *symbiosis* or *osmosis* for biology and *energy conservation* in physics. We presented the problems in a kind of card-sorting task to assess application of knowledge. Card-sorting tasks have been used in a variety of subject-specific studies as a time-economic way to assess relevant parts of a student's problem-solving ability (cf. Friege & Lind, 2006; Nehm & Ridgeway, 2011; Chi, Feltovich, & Glaser, 1981). In the card-sorting task, students were asked to sort the problems and name the underlying subject-specific problem schemes but not to solve them. The problem schemes were not given to the students, so they had to generate them themselves. For our 12 tasks, we adapted four different subject-specific problem schemes. Each problem scheme was suitable to deal with three different tasks. A point was awarded when the students named an adequate problem scheme and sorted the right tasks to it. These types of sorting tasks are

highly correlated with the actual solving of the problems (Friege & Lind, 2006; Nehm & Ridgeway, 2011).

## Validity of test score interpretation

To prove for the trustworthiness of our conclusions, we collected theoretical arguments and empirical evidence for different aspects of validity (Messick, 1987). We assured for content validity by adapting the content of the first year for our tests and by expert revision (cf. the "Test content regarding the curriculum" section). This expert revision ensured that test content is representative for different types of biological or physics knowledge. The use of established assessment methods (e.g. concept maps) for the assessment of each knowledge type is an argument for cognitive validity. Since sorting tasks are not an established measure for application of knowledge, we conducted a more extensive validation study for both tests. An Interpretation-Use-Argumentation (cf. Kane, 2013) supports the assumption that the test scores from both tests can be interpreted as objective, reliable, and valid measures for subject-specific problem-solving processes associated with application of knowledge (Binder, Schmiemann, & Theyßen, 2019). Construct validity is, among others, supported by the relatively low latent correlations between the knowledge types within each subject (cf. the "Quality of the prior knowledge tests" section). These medium and high latent correlations between the knowledge types indicate that our tests assess knowledge, even though in different types.

All arguments and evidence support the hypothesis that our test scores can validly be interpreted as measures of the knowledge types we wanted to assess.

## HS GPA

We assessed the HS GPA of the biology and physics freshmen using self-report. The HS GPA in Germany is rated in grades on a scale from 4.0 (pass) to 1.0 (best possible). We thus inverted the scale for regression analyses.

## Academic achievement

Academic achievement is often defined as a student's grades in certain courses at university or as a cumulative measure (Grade Point Average or Pass/Fail-Information) (e.g. Legg et al., 2001; Tai, Sadler, & Mintzes, 2006). For our analysis, we defined academic achievement as a conglomeration of success in the subject-specific exams of the first year. Because prior knowledge tests can only be valid predictors for subject-specific performance, we constrain academic achievement to the exams in the major. To form a variable for academic achievement, we define students who passed all those exams as successful and students who failed at least one exam as non-successful students. Students who skipped one exam in their first semester were included and coded as successful when they passed the same exam in their second semester.

Otherwise, they were coded as non-successful students. Students who did not take any exam were excluded from the study.

## Sample

The whole sample comprises 268 undergraduates enrolled in biology and physics at two German universities, of whom 162 students were freshmen in biology and 106 were freshmen in physics. In biology, 65 % were female and two persons do not identify as male or female. In physics, 76 % were male and one person did not identify as male or female. The ages ranged between 18 and 67 years ($Md = 20$) for physics students and between 17 and 52 years ($Md = 21$) for biology students.

## Analyses

We rated each item of each test with a coding scheme. For biology and physics, each test is associated with one specific prior knowledge type and was rated dichotomously or polytomously as described above. To ensure objectivity, 20% of the items (except knowledge of facts) were coded independently by two trained raters based on the coding schemes and Cohen's kappa (Cohen, 1960) was calculated as a measure of interrater-reliability.

The biology and physics assessments were scaled separately. In order to assess the four prior knowledge types, each test set was scaled following an item response theory (IRT) approach (Wilson, 2004). Because of the polytomous rating of some items, we applied a multidimensional Partial Credit Model to our data, an extension of the Rasch Model for polytomous items (Masters, 1982). The IRT approach has the advantage that the item fit can be evaluated for each item separately. In addition, IRT models separate item difficulty and person ability and allow for their separate evaluation (e.g. Bond & Fox, 2012; Boone & Scantlebury, 2006). To account for reliability, we calculated Warm's weighted likelihood estimate (WLE) reliability indices for the person parameters. The estimates refer to the variance of the person ability estimates calculated in the model and the average of the squared errors of these estimates (Adams, 2005). Since reliability measures are not considered the most important indicators of test quality (Adams, 2005), we used other indicators as well. Item infit statistics were evaluated for the items (Boone, Townsend, & Staver, 2011). This statistic shows how well an item fits the assumed construct that is to be measured and contributes to the validity of the prior knowledge tests. Third, we analysed the threshold parameters for polytomous items using Thurstonian thresholds (Bond & Fox, 2012). When threshold parameters are in an order, the different steps in an item are meaningful with regard to difficulty.

After we accounted for reliability, we computed WLE as a measure of person ability and examined the reliability of these measures and the latent correlations between the four prior knowledge types. WLE scores were used for further analysis in the regression models (Warm, 1989). We conducted statistical analyses using the software package R (R Core Team, 2014) with the package TAM (Test Analysis Modules) and IBM SPSS statistics 24.

The two most commonly used regression methods in prediction studies in science education concerning academic achievement are logistic and linear regression (e.g. Alzen, Langdon & Otero, 2018; Legg et al., 2001; Sadler & Tai, 2001). In our study, we applied binary logistic regression, in order to identify knowledge types, which contribute to overall academic achievement in the first year. We used the scoring of academic achievement as the dependent variable. Since our work is exploratory in nature, we strove for parsimony when building the regression models (Field, 2018). Therefore, we fitted a model with all knowledge types as predictors first and then excluded predictors without explanatory benefit step by step. All predictors were standardised before they were entered in the prediction model. To find significant models, we examined the likelihood ratio statistics of the baseline model and our regression models by chi-square tests. We utilised the Wald statistics to exclude knowledge types as predictors from the regression models (Field, 2018). Odds ratios were used to interpret the strength of the relationship between the knowledge types and academic achievement. They indicate the constant effect of the predictors on the likelihood of having academic achievement. Values of the odds ratio greater than 1 mean that as the predictor variable increases, so do the odds of having academic achievement (e.g. Alzen et al., 2018; Lehtamo, Juuti, Inkinen & Lavonen, 2018).

## Results

### Quality of the prior knowledge tests

As a precondition for further analyses, we first analysed the objectivity of the coding and reliability of the test measures (cf. the "Analyses" section).

Regarding objectivity, we found the interrater reliability to be substantial for most rated tests items (biology, $\kappa \geq .61$; physics, $\kappa \geq .64$). One item of the physics knowledge of meaning test showed only moderate ($\kappa = .44$) interrater reliability (Landis & Koch, 1977).

The WLE reliability of the test scores in biology is between 0.60 and 0.71, and in physics between 0.68 and 0.77. All reliability measures are acceptable (Bond & Fox, 2012). The standardised infit measures for the prior knowledge tests range from 0.88 to 1.12 in biology and from 0.73 to 1.21 in physics. As these fit measures do not exceed the cutoff values of 0.7–1.3, we assume that

the statistical model matches the empirical data (Bond & Fox, 2012). In both subjects, the threshold parameters for the knowledge of meaning tasks were found to be ordered and increasing in difficulty. Latent correlations between the knowledge types ranged between $.462 < r_{lat} < .663$ in biology and $.517 < r_{lat} < .757$ in physics. More information concerning reliability can be found in Additional file 1.

### HS GPA and academic achievement

In the first step, we utilised HS GPA as predictor of academic achievement. Full data sets with all outcome measures were available for 120 biology and 73 physics students.

For the biology sample, we found a non-significant model (A) in comparison to the baseline model ($\chi^2 = 1.11$, $p = .292$, $df = 1$). HS GPA made no significant contribution to the model (Table 2).

For the physics sample, the best fitting model (A) in comparison to the baseline model ($\chi^2 = 17.29$, $p < .001$, $df = 1$) includes the HS GPA as significant predictor for academic achievement (Table 3).

The physics prediction model (A) is able to predict 71.2% of the outcome variable.

### Knowledge types and academic achievement

In the second step, we used the knowledge types as predictors for academic achievement. Therefore, we added the prior knowledge types in the prediction models achieved in the first analysis of HS GPA.

For the biology sample, HS GPA was no longer included in the model because it made no significant contribution to the prediction of academic achievement. The best-fitting model for the biology sample, model E ($\chi^2 = 15.74$, $p < .001$, $df = 1$), has one significant predictor, knowledge of meaning (Table 2). Model E (Fig. 2a) classifies 68.3% of the biology students correctly. The odds ratio (OR) indicates that a rise of one logit in knowledge of meaning in biology raises the odds of being successful by a factor of 2.32 (95% $CI = 1.46$, 3.67; $p < .001$).

For the prediction of academic achievement with the prior knowledge types in physics, we included HS GPA in every prediction model because it was significant in the first model. The best-fitting model for the physics sample, model D ($\chi^2 = 41.56$, $p < .001$, $df = 3$), has two significant predictors, knowledge of meaning ($OR = 4.36$; 95% $CI = 1.72$, 11.13; $p = .002$) and application of knowledge ($OR = 4.73$; 95% $CI = 1.59$, 14.02; $p = .005$) (Table 3). HS GPA ($OR = 1.77$; 95% $CI = 0.69$, 4.56; $p = .234$) does not contribute to the model significantly if the knowledge types are included. Model D (Fig. 2b) predicts 82.2% of the cases correctly.

**Table 2** Summary of the logistic regression analysis for academic achievement in biology

|  | Model A | | Model B | | Model C | | Model D | | Model E | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | B (SE) | OR [95% CI] | B (SE) | OR [95% CI] | B (SE) | OR [95% CI] | B (SE) | OR [95% CI] | B (SE) | OR [95% CI] |
| Intercept | 0.769 (0.197) | 2.16 | 0.869 (0.219) | 2.38 | 0.842 (0.218) | 2.32 | 0.855 (0.216) | 2.35 | 0.867 (0.215) | 2.38 |
| HS GPA | 0.203 (0.193) | 1.23 [0.8, 1.8] | – | – | – | – | – | – | – | – |
| Knowledge of facts | – | – | 0.304 (0.261) | 1.36 [0.8, 2.3] | 0.292 (0.257) | 1.34 [0.8, 2.2] | 0.278 (0.256) | 1.32 [0.8, 2.2] | – | – |
| Knowledge of meaning | – | – | 0.780 (0.297) | 2.18** [1.2, 3.9] | 0.752 (0.282) | 2.12** [1.2, 3.7] | 0.706 (0.264) | 2.03** [1.2, 3.4] | 0.839 (0.235) | 2.32*** [1.5, 3.7] |
| Integration of knowledge | – | – | − 0.083 (0.270) | 0.921 [0.5, 1.6] | – | – | – | – | – | – |
| Application of knowledge | – | – | − 0.090 (0.230) | 0.914 [0.6, 1.4] | − 0.108 (0.222) | 0.90 [0.6, 1.4] | – | – | – | – |
| Correctly classified cases | 68.3% | | 70.0 % | | 70.8 % | | 70.8 % | | 68.3 % | |
| Nadelkerke's $R^2$ | .013 | | .188 | | .187 | | .185 | | .172 | |
| Chi-square tests | 1.11 n.s. | | 17.28** | | 17.18** | | 16.95*** | | 15.74*** | |

Chi-square tests test the deviance of the fitted model against the null model. Not significant predictors were excluded in the following model. Dependent variable = academic achievement (1 = passed all subject-specific courses)

*OR* odds ratio

**$p < .01$

***$p < .01$

## Discussion

### HS GPA and academic achievement

To answer Research Question I, "In how far does the HS GPA predict the likelihood of having academic achievement in the first year at university for biology and physics majors?", we examined the relationship between HS GPA and academic achievement at the end of the first year.

In our study, the HS GPA was not a significant predictor for academic achievement in biology in the first year at university. One reason for our finding could be that content covered in the first-year courses is disparate to the content

**Table 3** Summary of the logistic regression analysis for academic achievement in physics

|  | Model A | | Model B | | Model C | | Model D | |
|---|---|---|---|---|---|---|---|---|
| Parameter | B (SE) | OR [95% CI] | B (SE) | OR [95% CI] | B (SE) | OR [95% CI] | B (SE) | OR [95% CI] |
| Intercept | − 0.852 (0.312) | 0.43 | − 1.958 (0.581) | 0.14 | − 1.816 (0.540) | 0.16 | − 1.775 (0.530) | 0.17 |
| HS GPA | 1.282 (0.365) | 3.60*** [1.8, 7.4] | 0.513 (0.534) | 1.67 [0.6, 4.8] | 0.477 (0.505) | 1.61 [0.6, 4.3] | − 0.450 (0.440) | 1.77 [0.7, 4.6] |
| Knowledge of facts | – | – | 0.704 (0.503) | 2.02 [0.8, 5.4] | 0.754 (0.496) | 2.11 [0.8, 5.6] | – | – |
| Knowledge of meaning | – | – | 1.035 (0.530) | 2.82‡ [1.0, 8.0] | 1.232 (0.507) | 3.43* [1.3, 9.3] | 1.475 (0.477) | 4.37** [1.7, 11.1] |
| Integration of knowledge | – | – | 0.556 (0.521) | 1.74 [0.6, 4.8] | – | – | – | – |
| Application of knowledge | – | – | 1.259 (0.592) | 3.52* [1.1, 11.2] | 1.300 (0.576) | 3.67* [1.2, 11.3] | 1.553 (0.555) | 4.73** [1.6, 14.0] |
| Correctly classified cases | 71.2% | | 84.9 % | | 82.2 % | | 82.2 % | |
| Nadelkerke's $R^2$ | .287 | | .627 | | .615 | | .590 | |
| Chi-square Tests | 17.29*** | | 45.15*** | | 43.98*** | | 41.55*** | |

Chi-square tests test the deviance of the fitted model against the null model. Not significant predictors were excluded in the following model. Dependent variable = academic achievement (1 = passed all subject-specific courses)
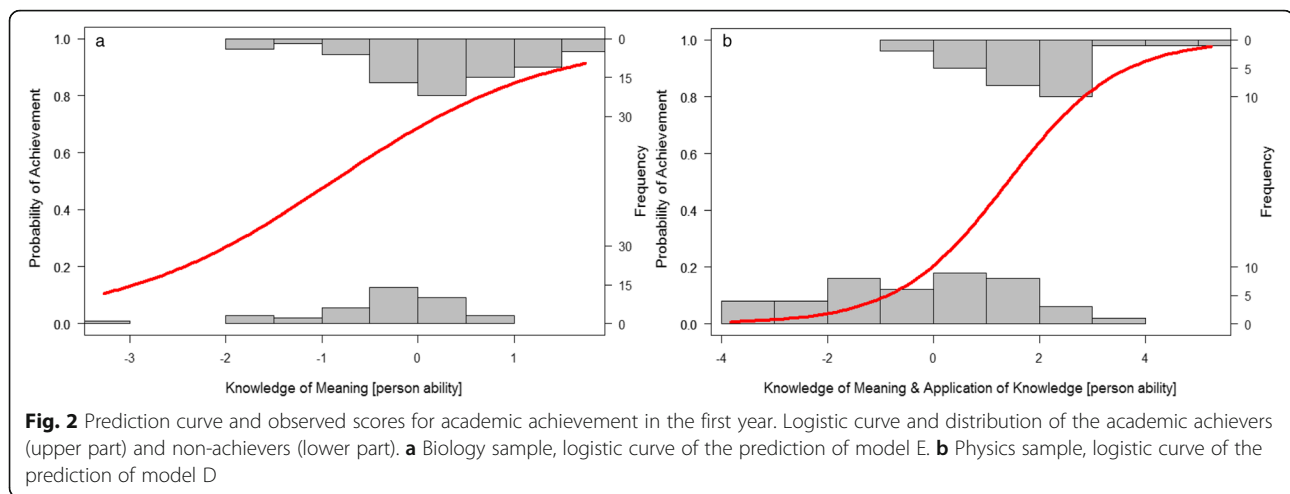
*OR* odds ratio

‡$p < .10$

*$p < .05$

**$p < .01$

***$p < .001$

**Fig. 2** Prediction curve and observed scores for academic achievement in the first year. Logistic curve and distribution of the academic achievers (upper part) and non-achievers (lower part). **a** Biology sample, logistic curve of the prediction of model E. **b** Physics sample, logistic curve of the prediction of model D

covered at school. High school biology content in Germany is limited to the topics of genetics, ecology, and evolution whereas the content at the university covers zoology, botany, and cellular biology. This explanation for our finding is supported by the findings of Loehr et al. (2012). Among several high school course grades, they identified only the grades of certain high school courses to be predictors of achievement in freshmen biology courses and the predictivity seemed to be an issue of content validity.

Another reason for our finding could be that the learning in the first year of biology is often limited to reproduction tasks. Therefore, higher cognitive skills, such as complex problem solving or reasoning that is related to the HS GPA, are not predictive for academic achievement here. In addition, the variance of the HS GPA of biology students is limited because it is a selection criterion for enrollment (*Numerus Clausus*).

For the physics students, we found the HS GPA as a significant predictor for achievement in the first year. A raise of one standard deviation in the HS GPA increases the odds of academic achievement more than threefold. This finding is in line with the findings of Sadler and Tai (2001), who found the HS GPA to be a significant predictor for the college grades in introductory courses. They also found that HS GPA ($B = 0.36$) is much less predictive than the high school course choice that is a more subject-specific indicator ($B = 2.26/3.51/4.32$ for regular/honours/AP). This is in accordance with our findings that the HS GPA is no longer a relevant predictor for academic achievement if the knowledge types are included in the model (Table 3).

The HS GPA, as a conglomeration of many different subjects, does not seem to be specific enough as a predictor for absolute achievement in the first year.

### Knowledge types and academic achievement

To answer Research Question II, "In how far are the knowledge types incrementally valid compared to the

HS GPA in predicting the likelihood of having academic achievement in the first year for biology and physics majors?", we examined the relationship between the four knowledge types and academic achievement at the end of the first year. The latent intercorrelations between the knowledge types support the hypothesis of four distinct knowledge types in both subjects. This is in line with the findings for mathematics by Hailikari et al. (2007) and for chemistry by Hailikari et al. (2008).

In biology, knowledge of meaning as a rather declarative knowledge type is the only knowledge type that predicts academic achievement in the first year. According to our results, we can state that an increase in knowledge about the concepts and principles of biology at the beginning of the introductory courses by one standard deviation ($SD_{KOM} = 0{,}88$ logits[1]), more than doubles the odds of being successful. Therefore, more prior knowledge about biological concepts and principles increases the odds of being successful in all subject-specific exams by the end of the first year at university. This finding differs from those of Hailikari and her colleagues, who always found procedural knowledge types as the best predictors in their studies (mathematics: Hailikari et al., 2007; pharmacy: Hailikari et al., 2008; chemistry: Hailikari & Nevgi, 2010). This stresses the subject-specific relevance of the prior knowledge types. Our result is related to the findings of Loehr et al. (2012) that high school courses that fostered deep understanding of biology concepts were positively related with academic achievement, and those of Lin, Liang and Tsai (2014), who found that students who contextualise learning as meaningful acquisition of knowledge are good academic achievers. A reason for our finding might be that

---

[1]WLE person abilities are estimated in logit units. Logits convert the raw scores (number of items solved correctly) into equal interval measures. For more information concerning logits and their advantages for science education research, see Boone and Scantlebury (2006).

students who memorise concepts are able to easily integrate new facts and ideas of biology with these concepts. In terms of knowledge integration theory, the declarative knowledge of rather low complexity seems necessary to construct new valid links among those and to other new science ideas (Liu et al., 2015).

To analyse why knowledge of meaning is the most important knowledge type for the prediction of academic achievement in biology, further research about the teaching in first-year classes would be required.

In our prediction models for physics, we controlled for the HS GPA. But as soon as the knowledge types were included, the HS GPA was no longer significant. This finding highlights the importance of subject-specific knowledge types as a prerequisite in physics. In physics, knowledge of meaning seems to play the same crucial role as a prerequisite for learning in introductory courses as it does in biology. Thereby, an increase in prior knowledge in knowledge of meaning by one standard deviation ($SD_{\text{KOM}} = 0.79$ logits), increases the odds of being successful in the introductory physics courses by a factor of four. In addition, application of knowledge is related to students' achievement. An increase of one standard deviation ($SD_{\text{AOK}} = 1.63$ logits) in the knowledge type quadruples the odds of being successful in the introductory physics courses. This means that students' knowledge of physics concepts as well as their ability to apply it and to solve subject-specific problems predict achievement in the first year. Therefore, more prior knowledge in these types of knowledge increases the odds of being successful in all exams in the first year in physics.

In contrast, Sadler and Tai (2001) found several indicators of problem-solving (e.g. homework done or the number of quantitative problems assigned) to be non-significant for college achievement in physics. In our study, we used a more direct measure of subject-specific problem-solving abilities, and our results suggest that problem solving is important for introductory physics courses at university. This finding is in line with other studies of problem-solving in physics (Friege & Lind, 2006).

## Conclusion

Our assessments for different types of prior knowledge in biology and physics can be used by lecturers to gain insight into students' knowledge at the beginning of or prior to biology or physics courses. The results of such assessments might be used for an early identification of at-risk students and reliable course guidance. Besides this immediate practical relevance for assessment and course guidance, the results are a profound starting point for further research and course development. Since our study is basically an analysis of correlations, further studies with interventions fostering selected knowledge types are needed to test for causality. The results of our study

suggest, which knowledge types should be selected for these intervention studies in each subject, i. e. knowledge of meaning in biology, knowledge of meaning, and application of knowledge in physics. Thus, our findings may guide the development of respective interventions and innovations. If the relationships between knowledge types and academic achievement turn out to be causal, a theory- and empirical-based optimization of pre-courses and introductory courses fostering relevant knowledge types is possible. The development of interventions fostering knowledge of meaning or application of knowledge can be based on current research findings. For example, Koretsky, Keeler, Ivanovitch, and Cao (2018) found that in introductory physics and biology courses, especially pedagogical tools, such as Audience Response Systems and Guided Inquiry Worksheets, provide the opportunity to improve students' knowledge regarding fundamental concepts, principles, and the application thereof. For the implementation of such practices, institutional conceptions of teaching practices (Lund & Stains, 2015) and adequate supporting measures for lecturers have to be taken into account (Bathgate et al., 2019).

Our findings are consistent with previous research on academic achievement in science subjects. Subject-specific prior knowledge is a good predictor of academic achievement in science classes (Loehr et al., 2012; Sadler & Tai, 2001). The current study contributes to relating different prior knowledge types to achievement in introductory courses in biology and physics. Our results complement previous results that were obtained based on the model of Hailikari et al. (2007), and they show that in biology, in contrast to other subjects, a declarative knowledge type is relevant for academic achievement. Furthermore, they add that the HS GPA is a good predictor in physics as it is a proxy for specific prior knowledge types.

The analysis of our study and the interpretation of our results entail some limitations. Our study is only able to predict academic achievement indicated by absolute success in all exams in the first year of biology and physics courses. This operationalisation of academic achievement is comparable over both subjects, but restrictive and somehow rough. It only distinguishes the successful students from other students. In addition to our operationalisation, academic achievement can be addressed by a variety of different indicators, such as grades at university, persistence to the sophomore year, length of time to degree or graduation, standardised exams, credit hours earned, or by more affective variables, such as satisfaction or well-being in the learning environment (Kuh, Kinzie, Buckley, Bridges, & Hayek, 2006). These indicators allow a more detailed view on academic achievement and should be addressed in further research. However, because our study aimed to compare biology and physics students, we had to recourse to this rough but comparable indicator.

The second research question might imply a direct comparison of biology and physics. But this is not possible since we used different tests in each subject, although we tried to keep the assessment in both subjects as similar as possible. Thus, the interpretation of the results should be restricted to the specific roles of the knowledge types within each subject.

## Additional file

**Additional file 1:** Supplementary Material. (PDF 314 kb)

### Abbreviations
ACORNS: Assessment of Contextual Reasoning on Natural Selection; ACT: American College Test; AP: Advanced Placement; CI: Confidence Interval; CINS: Concept Inventory on Natural Selection; FCI: Force Concept Inventory; GCA: Genetics Concept Assessment; HS GPA: High School Grade Point Average; IRT: Item Response Theory; OR: Odds Ratio; TAM: Test Analysis Modules; WLE: Warm's Weighted Likelihood Estimates; $\kappa$: Cohen's Kappa

### Authors' contributions
PS, AS, BS, and HT conceived the design of the study and coordinated the inquiries. All authors significantly contributed to the test development. TB performed the statistical analyses, drafted the manuscript, and participated in its design and coordination. TB, HT, and PS were major contributors in writing the manuscript. All authors contributed to the interpretation of the results and read and approved the final manuscript.

### Availability of data and materials
The datasets used and analysed during the current study are available from the corresponding author, TB, on reasonable request.

### Compliance with ethical standards
Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the DFG (Deutsche Forschungsgemeinschaft). All students participated in this study voluntarily but also received a small financial compensation.

### Competing interests
The authors declare that they have no conflicts of interest.

### Author details
[1]Biology Education, University of Duisburg-Essen, Universitaetsstr. 5, D-45117 Essen, Germany. [2]Aquatic Ecology, University of Duisburg-Essen, Universitaetsstr. 5, D-45117 Essen, Germany. [3]Leibniz University Hannover, Welfengarten 1A, D-30167 Hannover, Germany. [4]Physics Education, University of Duisburg-Essen, Universitaetsstr. 2, D-45117 Essen, Germany.

### References

Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation, 31*, 162–172.

Adelman, C. (1999). *Answers in the tool box: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment*. DC. ED Pubs: Washington.

Alters, B. J. (1995). Counseling physics students: A research basis. *The Physics Teacher, 33*, 413–415.

Alting, A., & Walser, A. D. (2007). *Retention and persistence of undergraduate engineering students: What happens after the first year? Proceedings of the American Society for Engineering Education Annual Conference and Exposition*. Hawaii: Honolulu https://peer.asee.org/2344.

Alzen, J. L., Langdon, L. S., & Otero, V. K. (2018). A logistic regression investigation of the relationship between the Learning Assistant model and failure rates in introductory STEM courses. *International journal of STEM education, 5*(1), 56. https://doi.org/10.1186/s40594-018-0152-1.

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of research in science teaching, 39*(10), 952–978.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369–406. https://doi.org/10.1037/0033-295X.89.4.369.

Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. New York: Grune and Stratton.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York, London: Holt, Rinehart and Winston.

Ausubel, D. P. (2000). *The acquisition and retention of knowledge: A cognitive view*. Dordrecht: Springer Netherlands.

Bathgate, M. E., Aragón, O. R., Cavanagh, A. J., Waterhouse, J. K., Frederick, J., & Graham, M. J. (2019). Perceived supports and evidence-based teaching in college STEM. *International journal of STEM education, 6*(1), 11. https://doi.org/10.1186/s40594-019-0166-3.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347–364.

Binder, T., Schmiemann, P., & Theyßen, H. (2019). Erfassung von fachspezifischen Problemlöseprozessen mit Sortieraufgaben in Biologie und Physik. [Assessing Subject-specific Problem Solving Processes Using Sorting Tasks in Biology and Physics] Zeitschrift für Didaktik der Naturwissenschaften. https://doi.org/10.1007/s40573-019-00090-x.

Birenbaum, M., & Dochy, F. (Eds.). (2012). Alternatives in assessment of achievements, learning processes and prior knowledge. (Vol. 42). Boston. Springer Science & Business Media.

Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.

Bond, T. G., & Fox, C. M. (2012). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). New York: Routledge Retrieved from http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10670521.

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education, 90*(2), 253–269.

Boone, W. J., Townsend, J. S., & Staver, J. (2011). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: An exemplar utilizing STEBI self-efficacy data. *Science Education, 95*, 258–280. https://doi.org/10.1002/sce.20413.

Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT® studies of classes graduating since 1980*. New York: The College Board.

Buschhüter, D., Spoden, C., Borowski, A. (2017). Studienerfolg im Physikstudium: Inkrementelle Validität physikalischen Fachwissens und physikalischer Kompetenz. Zeitschrift für Didaktik der Naturwissenschaften 23 (1):127–141.

Chang, M. J., Cerna, O., Han, J., & Saenz, V. (2008). The contradictory roles of institutional status in retaining underrepresented minorities in biomedical and behavioral science majors. *The Review of Higher Education, 31*, 433–464.

Chen, X., & Soldner, M. (2013). *STEM attrition: College students' paths into and out of STEM fields: Statistical analysis report (NCES 2014-001). National Center for Education Statistics, Institute of Education Sciences, U.S.* Washington, DC: Department of Education.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Chi, M. T. H., & Ceci, S. J. (1987). Content knowledge: Its role, representation, and restructuring in memory development. In H. W. Reese (Ed.), *Advances in child*

*development and behavior* (Vol. 20, pp. 91–142). New York: Academic Press. https://doi.org/10.1016/S0065-2407(08)60401-2.

Chingos, M. M. (2018). What matters most for college completion? Academic preparation is a key predictor of success. *AEI Paper & Studies, 3A*.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement., 20*, 37–46.

Core Team, R. (2014). R: A language and environment for statistical computing. In *R Foundation for Statistical Computing*. Vienna: Austria. URL http://www.R-project.org/.

de Jong, T., & Ferguson-Hessler, M. G. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*, 105–113.

Dochy, F. (1992). *Assessment of prior knowledge as a determinant for future learning*. Utrecht: Uitgeverij Lemma B.V.

Dochy, F. (1994). Prior knowledge and learning. In T. Husen & N. Postlewaithe (Eds.), *International encyclopedia of education* (2nd ed., pp. 4698–4702). London/New York: Pergamon.

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69*, 145–186.

Docktor, J., & Heller, K. (2008, October). Gender differences in both force concept inventory and introductory physics performance. In AIP Conference Proceedings (Vol. 1064, No. 1, pp. 15-18). AIP.

Eurostat (2016). Tertiary education statistics. http://ec.europa.eu/eurostat/statistics-explained/index.php/Tertiary_education_statistics#Further_Eurostat_information. Accessed 18 Juno 2018

Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics. SAGE edge.

Friege, G., & Lind, G. (2006). Types and qualities of knowledge and their relations to problem solving in physics. *International Journal of Science and Mathematics Education, 4*(3), 437–465.

Geller, C., Neumann, K., Boone, W. J., & Fischer, H. E. (2014). What makes the Finnish different in science? Assessing and comparing students' science learning in three countries. *International Journal of Science Education, 36*, 3042–3066.

Hailikari, T., Katajavuori, N., & Lindblom-Ylänne, S. (2008). The relevance of prior knowledge in learning and instructional design. *American Journal of Pharmaceutical Education, 72*, 113.

Hailikari, T., Nevgi, A., & Lindblom-Ylänne, S. (2007). Exploring alternative ways of assessing prior knowledge, its components and their relationship to achievement: A mathematics based case study. *Studies in Educational Evaluation, 33*, 320–337. https://doi.org/10.1016/j.stueduc.2007.07.007.

Hailikari, T. K., & Nevgi, A. (2010). How to diagnose at-risk students in chemistry: The case of prior knowledge assessment. *International Journal of Science Education, 32*, 2079–2095. https://doi.org/10.1080/09500690903369654.

Halloun, I. A., & Hestenes, D. (1985). The initial knowledge of college physics students. *American Journal of Physics, 53*, 1043–1055.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., & Elliot, A. J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology, 94*, 562.

Hart, G. E., & Cottle, P. D. (1993). Academic backgrounds and achievement in college physics. *The Physics Teacher, 31*, 470–475.

Hazari, Z., Tai, R. H., & Sadler, P. M. (2007). Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors. *Science Education, 91*(6), 847–876.

Hell, B., Trapmann, S., & Schuler, H. (2007). Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum. [A metaanalysis regarding the validity of academic aptitude tests in Germany]. *Empirische Pädagogik, 21*, 251–270.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher, 30*(3), 141–158.

Jonassen, D. H., Yacci, M., & Beissner, K. (2012). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Hoboken: Taylor and Francis Retrieved from http://gbv.eblib.com/patron/FullRecord.aspx?p=1099216.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73.

Kappe, R., & van der Flier, H. (2012). Predicting academic success in higher education: What's more important than being smart? *European Journal of Psychology of Education, 27*, 605–619.

Koretsky, M., Keeler, J., Ivanovitch, J., & Cao, Y. (2018). The role of pedagogical tools in active learning: A case for sense-making. *International journal of STEM education, 5*(1), 18. https://doi.org/10.1186/s40594-018-0116-5.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*, 212–218. https://doi.org/10.1207/s15430421tip4104_2.

Kuh, G. D., Kinzie, J. L., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). *What matters to student success: A review of the literature* (Vol. 8). Washington, DC: National Postsecondary Education Cooperative.

Kuncel, N. R., Hazlett, S. A., & Ones, D. R. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*, 162–181.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159–174.

Lee, U. J., Sbeglia, G. C., Ha, M., Finch, S. J., & Nehm, R. H. (2015). Clicker score trajectories and concept inventory scores as predictors for early warning systems for large STEM classes. *Journal of Science Education and Technology, 24*, 848–860. https://doi.org/10.1007/s10956-015-9568-2.

Legg, M. J., Legg, J. C., & Greenbowe, T. J. (2001). Analysis of success in general chemistry based on diagnostic testing using logistic regression. *Journal of Chemical Education, 78*, 1117–1121.

Lehtamo, S., Juuti, K., Inkinen, J., & Lavonen, J. (2018). Connection between academic emotions in situ and retention in the physics track: applying experience sampling method. *International journal of STEM education, 5*(1), 25. https://doi.org/10.1186/s40594-018-0126-3.

Lin, T.-C., Liang, J.-C., & Tsai, C.-C. (2014). Conceptions of memorizing and understanding in learning, and self-efficacy held by university biology majors. *International Journal of Science Education, 37*, 446–468. https://doi.org/10.1080/09500693.2014.992057.

Liu, O. L., Ryoo, K., Linn, M. C., Sato, E., & Svihla, V. (2015). Measuring knowledge integration learning of energy topics: A two-year longitudinal study. *International Journal of Science Education, 37*, 1044–1066.

Loehr, J. F., Almarode, J. T., Tai, R. H., & Sadler, P. M. (2012). High school and college biology: A multi-level model of the effects of high school courses on introductory course performance. *Journal of Biological Education, 46*, 165–172. https://doi.org/10.1080/00219266.2011.617767.

Lund, T. J., & Stains, M. (2015). The importance of context: an exploration of factors influencing the adoption of student-centered teaching among chemistry, biology, and physics faculty. *International Journal of STEM Education, 2*(1), 13. https://doi.org/10.1186/s40594-015-0026-8.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into Practice, 41*, 226–232.

Messick, S. (1987). Validity. ETS Research Report Series, 1987(2), i-208.

Mitchell, A. A., & Chi, M. T. (1984). Measuring knowledge within a domain. In P. Nagy (E.d.), The Representation of Cognitive Structure, 85–109.Toronto, Canada: Ontario Institute for Studies in Education.

Müller, J., Fischer, H. E., Borowski, A., & Lorke, A. (2017). Physikalisch-Mathematische Modellierung und Studienerfolg. [Physico-mathematical modelling and study success]. Implementation Fachdidaktischer Innovation Im Spiegel Von Forschung Und Praxis, 75.

Nehm, R. H., Beggrow, E., Opfer, J., & Ha, M. (2012). Reasoning about natural selection: Diagnosing contextual competency using the ACORNS instrument. *The American Biology Teacher, 74*(2), 92–98.

Nehm, R. H., & Ridgway, J. (2011). What do experts and novices "see" in evolutionary problems? *Evolution: Education and Outreach, 4*(4), 666.

Novak, J. D., & Gowin, B. (1999). *Learning how to learn*. Cambridge: Cambridge University Press.

Olson, S., & Riordan, D. G. (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Retrieved from https://files.eric.ed.gov/fulltext/ED541511.pdf

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (2001). *Using achievement tests/SATII: subject tests to demonstrate college and predict college grades: Sex, language, ethnic and parental education groups*. New York: The College Board.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes: A meta-analysis. *Psychological Bulletin, 130*, 261.

Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching, 33*, 569–600.

Sadler, P. M., & Tai, R. H. (2001). Success in introductory college physics: The role of high school preparation. *Science Education, 85*, 111–136.

Sadler, P. M., & Tai, R. H. (2007). Advanced placement exam scores as a predictor of performance in introductory college biology, chemistry and physics courses. *Science Educator, 16*, 1–19.

Schachtschneider, Y. (2016). *Studieneingangsvoraussetzungen und Studienerfolg im Fach Biologie [Freshmen requirements and academic success in biology]*. Berlin: Logos Verlag Berlin GmbH.

Schiefele, U., Krapp, A., & Winteler, A. (2014). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi, A. Krapp, & A. Renninger (Eds.), *The Role of interest in learning and development* (pp. 183–212). Hoboken: Taylor and Francis.

Schneider, W., & Pressley, M. (1997). *Memory development between 2 and 20*. New York: Springer.

Smith, M. K., Wood, W. B., & Knight, J. K. (2008). The genetics concept assessment: a new concept inventory for gauging student understanding of genetics. *CBE—Life Sciences Education, 7*(4), 422–430.

Sorge, S., Petersen, S., & Neumann, K. (2016). Die Bedeutung der Studierfähigkeit für den Studienerfolg im 1. Semester in Physik. [The relevance of the study capacity for success in introductory physics courses]. *Zeitschrift für Didaktik der Naturwissenschaften, 22*(1), 165–180.

Tai, R. H., Sadler, P. M., & Mintzes, J. J. (2006). Factors influencing college science success. *Journal of College Science Teaching, 36*, 52–66.

Thompson, R. A., & Zamboanga, B. L. (2003). Prior knowledge and its relevance to student achievement in introduction to psychology. *Teaching of Psychology, 30*, 96–101.

Van Riesen, S., Gijlers, H., Anjewierden, A., & de Jong, T. (2018). The influence of prior knowledge on experiment design guidance in a science inquiry context. *International Journal of Science Education, 40*, 1327–1344. https://doi.org/10.1080/09500693.2018.1477263.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*, 427–450. https://doi.org/10.1007/BF02294627.

Whalen, D. F., & Shelley, M. C. (2010). Academic success for STEM and non-STEM majors. *Journal of STEM Education: Innovations and Research, 11*, 45–60.

Wilson, M. (2004). *Constructing measures: An item response modeling approach*. Routledge.

Yin, Y., Vanides, J., Ruiz-Primo, M. A., Ayala, C. C., & Shavelson, R. J. (2005). Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use. *Journal of Research in Science Teaching, 42*, 166–184.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.