2014

Conceptions of giftedness and expertise put to the empirical test

Bettina Harder
University of Erlangen-Nuremberg

Wilma Vialle
University of Wollongong, wvialle@uow.edu.au

Albert Ziegler
University of Erlangen-Nuremberg

Conceptions of giftedness and expertise put to the empirical test

Abstract

Recent handbooks of giftedness or expertise propose a plethora of conceptions on the development of excellent performance but, to our knowledge, there are no comparative studies that provide empirical evidence of their validity to guide researchers and practitioners in their adoption of a particular conception. This study sought to close that gap by conducting an empirical comparison of the major approaches to giftedness and expertise currently in use: the IQ model, the performance model, the moderator model, and the systemic model. The four models were tested in a longitudinal study with a sample of N=350 German students attending university preparatory schools; 25% of the sample had been assigned to special classes for the gifted. The construct and predictive validity of the four models were tested by means of structural equation modeling. Theoretical considerations along with our results indicated a differentiation among the models whereby some could only predict while others could also explain the emergence of excellent performance and thereby yield valuable information for the design of interventions. The empirical comparison of the approaches showed that they were unequally suited for the two challenges. For prediction purposes, the performance approach proved best while, for explanations, the moderator and systemic approaches were the most promising candidates. Even so, the latter did demonstrate conceptual and/or methodological problems. The IQ approach was superseded by the other approaches on both prediction and explanation. Implications and limitations of the findings are discussed.

Keywords

empirical, test, put, giftedness, expertise, conceptions

Disciplines

Education | Social and Behavioral Sciences

Publication Details

Harder, B., Vialle, W. & Ziegler, A. (2014). Conceptions of giftedness and expertise put to the empirical test. High Ability Studies, 25 (2), 83-120.

Conceptions of giftedness and expertise put to the empirical test

Bettina Harder[1], Wilma Vialle[2], & Albert Ziegler[1]

[1] Chair of Educational Psychology and Excellence Research, University of Erlangen-Nuremberg, Germany

[2] Faculty of Social Sciences, University of Wollongong, Australia

Corresponding author:

Bettina Harder, PhD

Chair of Educational Psychology and Excellence Research

University of Erlangen-Nuremberg

90478 Nuremberg

Germany

Phone: 0049 (0)911 5302-598

Fax: 0049 (0)911 5302-166

Email: bettina.harder@fau.de

Revised resubmission, Sept. 15th 2014

Abstract

Recent handbooks of giftedness or expertise propose a plethora of conceptions on the development of excellent performance but, to our knowledge, there are no comparative studies that provide empirical evidence of their validity to guide researchers and practitioners in their adoption of a particular conception. This study sought to close that gap by conducting an empirical comparison of the major approaches to giftedness and expertise currently in use: the IQ-model, the performance model, the moderator model, and the systemic model. The four models were tested in a longitudinal study with a sample of N = 350 German students attending university preparatory schools; 25% of the sample had been assigned to special classes for the gifted. The construct and predictive validity of the four models were tested by means of structural equation modeling. Theoretical considerations along with our results indicated a differentiation among the models whereby some could only *predict* while others could also *explain* the emergence of excellent performance and thereby yield valuable information for the design of interventions. The empirical comparison of the approaches showed that they were unequally suited for the two challenges. For prediction purposes, the performance approach proved best while, for explanations, the moderator and systemic approaches were the most promising candidates. Even so, the latter did demonstrate conceptual and/or methodological problems. The IQ-approach was superseded by the other approaches on both prediction and explanation. Implications and limitations of the findings are discussed.

Keywords: conceptions of giftedness, expertise, empirical comparison, intelligence, performance, moderators, systemic

Conceptions of giftedness and expertise put to the empirical test

Creating a Reasoned Basis for Theory Choice

The field of giftedness research currently faces some theoretical and methodological challenges (for a detailed overview, see Harder, 2012c). One of the major problems is that extant theories of giftedness and expertise research differ greatly in their explanations of outstanding performance while we lack a reasoned basis for deciding which theory to apply to our research or practical questions (cf. Ziegler, 2005). This paper is a first step towards building an empirical foundation for rational decisions for or against a theory, by providing an empirical comparison of current theories explaining outstanding performance.

Recent handbooks of giftedness present several conceptions of giftedness but no comparative empirical studies (Heller, Mönks, Sternberg, & Subotnik, 2002; Sternberg & Davidson, 2005). While there is some theoretical underpinning and empirical evidence for all of these conceptions — as the authors proposing the conception usually outline — the quality of the validating studies varies greatly and is often flawed (cf. Harder, 2012c; Ziegler & Heller, 2002). Amidst such competing conceptions, we need to know which theory is best suited to guide our research and practice instead of basing our decisions on personal taste and selective corroborative evidence.

Demands on Theories of Giftedness and Expertise

A good theory needs to fulfil the demands of researchers and practitioners in the field of giftedness and expertise. In research, we first and foremost want to *explain* the emergence of outstanding performance, that is, to identify the causal factors in the developmental process. Understanding the causalities in this process would provide starting points for the design and implementation of targeted interventions to optimize children’s and adults’ development. As this theoretical groundwork lays the basis for practical implementations, it is of enormous importance to create a realistic and yet manageable theory so that efforts in identification and individual programming are not misguided. Further, educational systems, institutions and governments seek ways to implement effective gifted education (Ziegler, Stoeger, Harder, & Balestrini, 2013) and, hence, urgently need good theories on which to build large programs. For these practical purposes, good *predictors* may sometimes be sufficient to make a decision but without *explanations* the undertaking may be “flying blind”.

To meet these requirements, theories of giftedness should be able to explain, or at least predict, future excellent achievements. Conceptions that will provide sound *explanations* need to fulfil all three kinds of validity. The first of these is content validity, that is, specifying all relevant aspects, which can only be ascertained by argumentation, not by empirical testing (for a thorough theoretical evaluation of the conceptions of giftedness, see Harder, 2012c). The second, construct validity, demands that the specific features of the conception (the constructs and their interaction), as theorized, withstand an empirical test. Finally, the third, empirical validity, can be further divided into associations of gifts and other criteria in the present and in the future. Predictions of future success are particularly interesting. For a theory to be only *predictive* (in contrast to be explanatory), it is sufficient to fulfil predictive empirical validity, that is, to be a stable correlate of future performance. The aim of this paper, then, is to unveil the construct and predictive validity of different conceptions of giftedness and expertise through empirical testing.

Criteria for an Empirical Test of Theory

To build a solid empirical basis for theory choice, the empirical comparison should fulfil high research standards. A literature review (cf. Harder, 2012c) revealed criteria regarding the content of studies and their methodology.

The criteria proposed for study content include:

- Proof of construct and predictive validity: empirical verification of the complete theory (all components, interactions), or at least evidence for the relevance of all proposed components; and,
- Exclusion of alternative explanations, that is, testing theories against each other to confirm one's predominance.

The criteria for study methodology comprise:

- Adherence to common research standards (e.g., study design, assessment instruments, etc.);
- Longitudinal field studies to evaluate the theory's validity over the natural developmental course of learners (Helmke & Weinert, 1997);
- Adequate methods of data analysis for modeling the complexity assumed in the theory, for example, structural equation modeling (SEM, Helmke & Weinert, 1997); and,
- Methods of data analysis accounting for hierarchical data structures (Helmke & Weinert, 1997).

Our study fulfils the demands to a large extent. Each approach to giftedness is modelled as completely as possible to test validities and to compare them, using data from one common sample and thereby providing highly comparable results. The methodological issues were met with the exception of hierarchical data analysis, which was impossible to carry out due to the high number of variables to be included in the models of giftedness (the number of variables modelled in hierarchical SEMs is limited to the number of clusters minus one, L. K. Muthén & Muthén, 1998-2009). The methodology of SEM[1] (cf. Kaplan, 2000; Tabachnick & Fidell, 2001) is best suited for the research question at hand because it allows

[1] SEM recalculates the empirical covariances found in the dataset from the model one implements with latent variables. As most of the variables of interest in educational research cannot be directly observed, latent variables are supposed to capture the essence, the common variance, of the different observed variables and represent the construct more validly (i.e., the measurement model or the confirmatory factor analyses in the model). Subsequently, regressions and correlations can be implemented between these latent factors (i.e., the structural model) to test the hypothesized influences and relations.

us to test theoretically-based relationships and interactions among variables such as the ones proposed in the conceptions of giftedness.

Approaches to Explain Extraordinary Performance

Extant theories of giftedness and expertise try to explain the genesis of extraordinary performances in various domains. Essentially, they consist of three main statements: theories of giftedness describe some sort of *potential*, which undergoes a process of *transformation* to result in excellent *performance*. The differences between specific theories can generally be traced back to differences in the understanding of these three core concepts and the weight they are assigned for the emergence of excellence (ranging from non-consideration to centrality). A closer look at existing theories of giftedness and achievement excellence with these three concepts in mind indicates that there are five different approaches. Drawing on the four-group classification suggested by Mönks and Katzko (2005), the five approaches can be characterized as the *psychometric single component, multi-component, performance based, moderator* and *systemic approach*.

The Psychometric Single Component Approach and the IQ Model

Early research on giftedness was marked by the advance of *psychometric single component conceptions*. These conceptualize giftedness as a stable personality trait (potential), which is deemed independent of other variables such as environmental or historical context. Under this tradition, in the field of academic performance 'intelligence' is a synonym for giftedness and viewed as domain-independent (e.g., Carroll, 1993; Horn & Cattell, 1966; Spearman, 1904). The main focus of the approach was the description and measurement of this trait, using instruments such as the intelligence tests utilized today. Psychometric single component theories do not differentiate between potential and performance, nor is a transformation specified. Potential or high intelligence automatically becomes evident in extraordinary performance. The underlying process is fundamentally autocatalytic, meaning that the transformation from potential to exceptional achievement is a matter of natural development.

For empirical testing this approach was represented by a classic IQ-model in which general intelligence is the sole contributor to and thereby predictor of extraordinary achievement (Rost, 2009)(cf. Figure 1).

The Multi-Component Approach

Due to the very restricted perspective of psychometric single-component theories, *multi-component theories* were introduced. These take an analytic approach to the cognitive processes of information processing, identifying different relevant components (e.g., Sternberg, 1985; Sternberg, 2005), and/or they consider other (non-cognitive) components of potential apart from intelligence. Examples of such additional components are creativity or task commitment (e.g., Mönks & Katzko, 2005; Renzulli, 1986, 2005). To define giftedness, all of the hypothesized components must emerge mutually as they are all assumed to be necessary contributions of equal importance to giftedness (potential). Their interaction is not made explicit nor is the transformation into performance specified. Consequently, potential and performance still cannot be clearly distinguished.

To empirically investigate the approaches, they had to be organized into specific, testable models, which entailed identifying a representative, but concise, variant of each approach. For the multi-component approach, this was impossible as the conceptions differ enormously and often provide non-operationalizable definitions of their constructs (e.g., the WICS-model by Sternberg, 2005). Hence, this approach did not form part of our empirical comparison.

The Performance Based Approach and the Performance Model

The shortcomings of the multi-component approach are remedied by *performance based theories*. Performance based conceptions are strongly influenced by expertise research and explicitly describe the transformation process. Therefore, they hold the key to possible interventions to foster the transformation of potential into performance. Performance based theories (e.g., Ericsson, Krampe, & Tesch-Römer, 1993) are highly dynamic in nature and focus on the learning process as the causal factor leading to ever higher levels of performance, thereby pushing personality traits into the background. Efficient learning is called deliberate practice, that is, learning that is consciously intended to improve one's skills and depends on three limiting factors, motivation, material means and dispensable effort. Potential is defined as the individual's current competence level in a specific domain that has been acquired through previous learning and which builds the basis for future knowledge acquisition.

Following this line of thought we modelled the performance approach by using former performance (inhering the individuals' capacities for deliberate practice) as the sole contributor to and predictor of later performance (cf. Figure 1). Assessing deliberate practice for school subjects in students as well as assessing additional undefined limiting factors is extremely hard and was therefore disregarded in our study.

PLACE FIGURE 1 ABOUT HERE

The Moderator Approach and Model

The process-oriented perspective of the performance based approach made way for the investigation of moderating influences on the transformation. The result of this extension of perspective was the further development of multi-component conceptions into *moderator conceptions*. Moderator theories assume that domain-specific gifts (basic skills, such as intelligence for academic domains) are transformed via learning into increasingly differentiated performances in that domain (cf. Figure 1). This process is moderated by various personal and environmental factors which imply their secondary role in the development of excellence. In essence, the differences compared to multi-component conceptions lie in the implementation of the transformation process, including the differentiation between potential and performance, and in the application of a hierarchy to the role of causal factors which all taken together build the potential for excellent achievement. The conceptions commonly include all thinkable gifts and moderating factors, assigned to the two large-scale categories of non-cognitive and environmental variables (e.g., Gagné, 2005; Heller, Perleth, & Lim, 2005). Examples for the first are motivational goal orientation, learning behavior, coping strategies for stress, and test anxiety while family climate, class climate, quality of instruction, attitude of peers, role models, and critical life events represent exemplary environmental influences. Beside that common ground, the theories differ in their assumptions of interactions among those factors (with Perleth’s theory assuming the highest level of interaction).

As moderator conceptions categorise various concepts under the term non-cognitive personality variables and environmental factors without clearly specifying the concepts to investigate, we decided to focus on the following areas in our study. We selected the broad concepts of motivation and learning behavior as central non-cognitive factors. *Motivation* was operationalized by academic self-concept, motivational goal orientations, and interest for the school subjects. Those concepts represent facets of the common expectancy-value-theories of motivation (e.g., Wigfield & Eccles, 2000). The academic self-concept was assessed in general and subject-related components (cf. Marsh, 1990); motivational goal orientations were measured according to the 2 x 2 theory by Cury et al. (2006), differentiating performance versus mastery and approach versus avoidance motivation; and, interest was conceptualized as the individual, relatively-persisting commitment to a topic (cf. Hidi, 1990; Renninger, 2000). *Learning behavior* was divided into goal commitment, effort, concentration, and working attitude, which represent two facets of situational motivation and two of volitional regulation of learning (Boekaerts, 1997) deemed important for self-regulated learning.

The proposed environmental factors of influence always cover the school- and family-related environments of students, which is why we focused on school and class climate and several family-related variables. *School and class climate* encompasses the aspects of perceived warmth, pressure exerted on students during lessons, student centeredness of teaching and the learning climate in class (cf. Eder & Mayr, 2000). Of the *family-related* environmental variables only the aspect of familial appreciation of school matters qualified for further analyses. The scale captures school- and learning-related attitudes of parents and siblings. Beside the choice of variables, which moderator theories leave to the investigator, we also had to decide on their interaction. We included all possible interactions so as not to impose restrictions on the tested model.

The Systemic Approach and Model

The fifth and last approach represents a further extension of the moderator theories' interaction idea towards a *systemic* perspective, which no longer perceives potential as a person's characteristic or sum of characteristics but as a property of the system that the person is part of (e.g., Dai & Renzulli, 2008; Ziegler, 2005). Hence performance cannot be explained analytically by means of single components but must be viewed from a holistic perspective. Performance is the result of complex interactions between individuals and their environment, which lead to an equally complex transformation process. The complete system with its current performance capacities (domain-specific system's potential) must evolve towards a more effective or more competent structure, which enables the person to demonstrate performance on a higher level in this domain.

Currently, only one theory provides the precision and comprehensiveness to qualify for empirical testing of the *systemic approach*. That theory is Ziegler's Actiotope model of giftedness (2005). Other systemic theories work with a general systemic understanding but do not specify the interacting factors leading to outstanding achievement. The best alternatives would be Dai and Renzulli's Snowflake model (2008), which describes alternating phases of reaching expertise but not the interplay of relevant factors, and Csikszentmihalyi's theory of creative development (2002), which can barely be applied to achievement in general. The Actiotope model focuses on actions performed by a learner in interaction with her/his environment and therefore uses a different terminology than trait based theories. It depicts four interacting components when it comes to performing, which are steered by five developmental mechanisms and depend on the general system state (see Figure 1). The *Actiotope components* are the environment, which interacts with the person where three functional entities are distinguished: first, the person's goals which, for example, give direction and make one initiate actions; second, the person's action repertoire ,which represents all theoretically possible actions for this person; third, the subjective action space where aspects of the inner and outer state are represented to enable decision making on which action to perform given environmental circumstances, personal goals and available action options in the action repertoire. The *developmental mechanisms* that are assumed to be at play to propel the level of actions performed in an Actiotope consist in the (1) ability to recognize when actions have been successful, (2) conditional knowledge on when best to implement which action, (3) ability to vary actions in order to fit the current situation, (4) anticipating of future obstacles and challenges, and (5) availability of high quality feedback. The interplay of the components and the developmental mechanisms depends on the general *system state* which should reconcile the attributes of modifiability and stability at the same time. Stability is important to assure the system's functioning, for example, a secure family situation or unfettered access to learning materials. Modifiability, on the other hand, is necessary to allow adaptations to higher levels of expertise, for example, the change of working attitudes or strategies to allow for more organized and effective learning. The highly dynamic and interactive nature of this systemic theory claims that all variables directly interact with each other. This was supported by prior analyses (Harder, 2012b, 2012c) and will be operationalized in the model test as far as possible with the available statistical methods.

Research Questions

This paper seeks to answer three research questions. Each approach to giftedness tries to explain the emergence of excellent achievement and thereby assumes a somehow causal relationship between the proposed antecedents of high performance and the achievement. Hence, they should all be valid in respect to construct and predictive validity, which is the first proposition to be investigated in this paper. The second question refers to possible changes in the validity of the approaches over the course of time and thereby the course of the learners' development. Hence, the validities were tested at two different points in time in an attempt to replicate the results. Third, we wanted to compare the different approaches to be able to tell which is best suited for (perhaps different) scientific and practical purposes.

Method

Study Design and Procedure

The longitudinal study comprised three measuring points. Students were assessed at the beginning of grade five (t1), that is, when they started secondary school; at the end of grade five (t2); and, at the end of grade six (t3). This design allowed us to replicate predictions from t1 to t2 within the second year (t2 to t3).

The studied variables were measured via tests and questionnaires for the sample students as well as questionnaires for their parents. For the students, each measuring point comprised two to four assessment appointments within a time interval of two to six weeks. Assessments took place during regular lessons at school and were assigned two or three lessons of 45 minutes duration. Students filled out all questionnaires and tests during these appointments. Parents' questionnaires were handed out to the students and collected by the teachers who returned them to the university. At t1, all variables were assessed; at t2, all but intelligence was assessed; and, at t3, only achievement measures were taken. Hence intelligence is the only variable that was assessed only once; all other predictors of achievement were measured at t1 and t2 while the dependent variable of achievement was measured at all times.

Participants

The sample consisted of N = 350 students of the German "Gymnasium" track (highest level, university preparatory, secondary school in the tracking system) and included students from special classes for the gifted (25%). The sample comprised students from six different schools with special classes for the gifted located in major cities in the south of Germany. To be admitted to special classes, students had to provide an (often independent) IQ-test result of at least 120 to the school. They were also likely to differ in other aspects, which is why sample characteristics are described for the complete sample and both subsamples (indexed G for gifted classes and R for regular classes).

At the beginning of the study (grade five), the mean age of the complete sample was 10.65 years (SD = 0.51) with gifted students being significantly younger than students from regular classes, $t(103.76) = -4.60$, $p < .001$, with the means being $M_G = 10.37$, $SD_G = 0.69$, and $M_R = 10.73$, $SD_R = 0.41$. The sample consisted of 60 % boys but the gifted and regular classes did not differ in this distribution, $t(348) = 0.062$, $p = .950$.

The complete sample showed an IQ of M = 110.29 and SD = 10.96 in the test conducted within this study. As can be expected by the IQ-based selection procedure for the gifted classes, the two types of classes differed significantly, $t(338) = 11.53$, $p < .001$, with $M_G = 120.9$, $SD_G = 9.1$ vs. $M_R = 107.4$, $SD_G = 9.4$. IQ-differences mirrored differences in socioeconomic background. The parents of students in the regular classes had lower levels of education compared to parents of the students in the gifted classes with $\chi^2(5) = 32.47$, $p < .000$ for fathers and $\chi^2(4) = 30.28$, $p < .000$ for mothers (categories: no school certificate, certificate of lowest level secondary school “Hauptschule”, mid-level “Realschule”, or highest level secondary school “Gymnasium”, diploma/bachelor/master degree, PhD).

Before predicting performances in detailed analyses, general differences in performance-levels needed to be considered. This was completed through the use of ANOVAs with the between-factor “group” and within-factor “time” for each subject (Greenhouse-Geisser corrected *df*s were used where necessary). Performance tests were taken in the subjects of German, English (first foreign language), and mathematics (for the following *F*-values, indexed by G, E, M respectively). Figure 2 displays the groups’ performance results across the three measuring points. In all three subjects, the gifted classes performed at a substantially higher level than the regular classes with $F_G(1, 286) = 51.83$, $p < .001$, $F_E(1, 261) = 57.25$, $p < .001$, and $F_M(1, 277) = 53.95$, $p < .001$. Both groups displayed significant development in performance levels over time with $F_G(2, 572) = 124.48$, $p < .001$, $F_E(1.86, 484.74) = 3624.53$, $p < .001$, $F_M(2, 554) = 120.35$, $p < .001$. Aside from the main effects, the analyses only revealed a significant interaction effect for mathematics with $F_M(2, 554) = 3.34$, $p < .05$, while for the other two subjects interaction effects did not reach the significance level with $F_G(2, 572) = 2.13$, $p = .120$, and $F_E(1.86, 484.74) = 1.77$, $p = .174$. To sum up, gifted classes showed higher performance levels in all investigated subjects and all students made progress in their knowledge and abilities of the subjects. In mathematics, gifted classes showed a steeper performance increase, that is, quicker progress between measuring points one and two, but then plateaued (see Figure 2) while the regular classes displayed a more linear increase.

PLACE FIGURE 2 ABOUT HERE

Measures

As all approaches try to explain excellent achievements, performance measures form a crucial part of every model. Furthermore, the constructs of intelligence, non-cognitive variables, environmental variables, and the Actiotope had to be assessed. All scales are described in the following section. Reliabilities are reported comprehensively in the form of Rasch-reliabilities and Rasch separation indices in Table 1.

Performance.

Scholastic performance was measured by objective tests in the subjects of German, English (the students' first foreign language learned), and mathematics. With the goal of assessing competences representative of the subjects, in *German*, reading comprehension was measured as one key competence targeted in the first years of Gymnasium. For t1 and t2, the FLVT 5-6 ("Frankfurter Leseverständnistest", Souvignier, Trenk-Hinterberger, Adam-Schwebe, & Gold, 2008) was used, which contains two texts with 18 multiple-choice comprehension questions for each. In a pilot-study, this test proved to be very easy for gifted students at t3. Thus, to avoid ceiling effects, the LESEN 8-9 ("Lesetestbatterie", Bäuerlein, Lenhard, & Schneider, 2012), a test designed for 8th to 9th graders with an identical test concept (two texts, each with 19 multiple-choice comprehension questions), was used for t3. Reliabilities were satisfactory with only one separation index slightly below the critical value (t2, person SEP = 1.85; see Table 1).

In *English*, to provide a representative performance measure, new tests were designed for the study (and pilot-tested) due to a lack of contemporary tests that covered the complete content of beginners' English classes. The English test (Harder & Ziegler, 2009) consisted of a vocabulary task, orthography task, grammar tasks, reading comprehension tasks, and two tasks testing phonetic knowledge (phoneme discrimination and syllable emphasis). All reliability measures were satisfactory (Table 1).

In *mathematics*, new tests were also designed (and pilot-tested) to assess the competences of high every-day relevance being promoted in the first years of Gymnasium. The tests (Weiß & Schneider, 2009) contained arithmetic problems as well as problems presented in text form. Reliabilities were satisfactory with only one low person separation index of SEP = 1.85 for t1 (Table 1).

Intelligence.

Students' intelligence was assessed at the beginning of grade five with the respective form of the KFT4-12+R ("Kognitiver Fähigkeitstest", German version of the cognitive abilities test for 4th to 12th grade and higher, revised edition, Heller & Perleth, 2000). The KFT4-12+R stands in accordance with several factor theories of intelligence, as the authors point out, using three subscales consisting of verbal, numeric and figural tasks. The person reliability for each subscale was satisfactory but the corresponding separation indices fell slightly below the critical value with SEP = 1.73 (numeric scale) to SEP = 1.99 (nonverbal scale). Item statistics were strong (Table 1).

Non-cognitive variables.

Motivation was operationalized by assessing academic self-concept, motivational goal orientations, and interest for each of the subjects (mathematics, German, and English). Academic self-concept was assessed using a translated and subject-specific version of the Self Description Questionnaire (Marsh, 1990) containing three items each on general academic, mathematics, English and German self-concepts, for example, "I have always been good in maths". The four different motivational goal orientations resulting from the 2 x 2 theory (Cury et al., 2006) were assessed by the translated and subject-specific adaptation of Elliot's Achievement Goal Questionnaire (1997). For example, for performance-avoidance goals an item is "In maths it is my goal not to perform worse than other students". Each goal orientation was measured with three items, leading to 12 items for each of the three subjects. To assess interest we used three items per subject taken from a national longitudinal mathematics study in Germany (PALMA, “Projekt zur Analyse der Leistungsentwicklung in Mathematik”, Pekrun, Götz, Zirngibl, vom Hofe, & Blum, 2002), for example, “I exert myself in English because I am interested in this subject”. All items were rated on a five-point Likert scale. Rasch-analyses revealed that motivational measures were homogeneous across subjects and, in the case of goal orientations, also across approach and avoidance tendencies. Therefore, the according items were summarized in combined scales. Reliability measures of those scales were quite good (Table 1) with one reliability and some separation indices below the critical values.

Learning behavior was divided into goal commitment, effort, concentration, and working attitude. The first three aspects were assessed with a questionnaire by Stumpf (2008) containing 10 items on effort (i.e., persistence in difficult learning situations) such as “I try to keep on doing my homework even if I don’t want to”, 15 items on goal commitment such as “I don’t give up in the face of obstacles”, and 14 items on concentration (abilities to prevent getting distracted during learning) such as “if necessary, I can concentrate for a long time on one topic”. All items were rated on a five-point Likert scale. Working attitude was measured with the LAVI (“Lern- und Arbeitsverhaltensinventar”, Inventory for learning and working behavior, Keller & Thiel, 1998). Thirty items present learning situations with three possible solutions to choose from: one good solution (3 points), one mixing up learning and other activities and leading to ineffective learning (2 points) and one which postpones the learning process (1 point). Reliabilities for the subscales were all satisfactory, with only one separation index below the critical value (Table 1).

Environmental variables.

Environmental variables covered the school and family related environments of students. *School and classroom climate* was measured by the LFSK (“Linzer Fragebogen zum Schul- und Klassenklima”, Linz questionnaire of school and class climate, Eder & Mayr, 2000), a student questionnaire consisting of 16 scales. According to the Rasch-analysis of the questionnaire data, only some scales were used for further analysis. These scales covered the topics of school climate (warmth, six items e.g., “My school is little/very supportive”) and class climate. Class climate is subdivided into three components with several subscales of three items each:

- Social and learning pressure with the subscales strictness of teachers, fairness of teachers, teachers’ practice of social comparisons of performance in the classroom, pressure to learn, and pressure during lessons such as leaving slower students behind. A sample item for strictness is “If someone is not working properly, teachers threaten him or her with a bad grade report card.”
- Student centeredness with the subscales pedagogic engagement of teachers, students’ right to have a say, quality of instruction, student participation, and control exerted by teachers like correcting homework. A sample item for pedagogic engagement is “Our teachers are honestly happy when they have taught us something.”
- Learning climate with the subscales sense of community, willingness to learn, rivalry, and disturbances. A sample item for rivalry is “If someone makes a mistake the others are gloating.”

All items were rated on a five-point Likert scale with the extremes “not true at all” and “very true”. Only the subscale, social warmth, presented contrary attributes as poles of the five-point rating scale. Reliabilities and separation indices were satisfactory with the exception of two separation indices below the value of 2 (see Table 1).

The *familial environment* was assessed with four items on the appreciation of school matters within the family (Preckel, 2008) with items such as “In our family school plays an important role” requiring a rating on a five-point Likert scale. The items could not be Rasch-scaled and were therefore modelled in the SEM as indicators on one latent factor.

Actiotope.

Students' Actiotopes were assessed with a 10-scale questionnaire (Ziegler, 2008a). The scales measured the Actiotope components (subjective action space, goals, environment), the developmental mechanisms (anticipation of future challenges, knowledge applicability, feedback they receive, ability to generate action variants, knowledge of correctness of actions), and the system aspects (modifiability, stability) with five items per scale. Each item, for example, "So far, I haven't been able to use outside school what I have learned at school", was rated on a five-point Likert scale. As the questionnaire cannot measure the action repertoire, that is, abilities, this further Actiotope component was covered by using the latent factor scores of the three school achievement tests. Reliabilities of the 10 subscales were very good except for the person reliability of the environment-scale. Separation indices were low for some of the Actiotope scales (Table 1).

PLACE TABLE 1 ABOUT HERE

Data Preparation and Analysis

Rasch-scaling and test-linking.

All test and questionnaire data were Rasch-scaled before being used for further analyses to ensure interval scale quality (cf., Bond & Fox, 2007). The scale of familial appreciation of school could not be Rasch-scaled and was therefore treated as data of ordinal rather than interval nature. All Rasch-analyses were carried out with Winsteps 3.71.0.1 (Linacre, 2011). Data consisted of dichotomous and Likert-scale responses which were scaled with the dichotomous (Rasch, 1960) resp. partial credit Rasch model (Masters, 1982).

Dichotomous test data were Rasch-scaled and then linked across the measuring points to provide valid performance measures along the two year development of students' knowledge in the face of changing items, that is, adapting the tests' difficulty to the higher performance levels. Performance tests were linked according to a common item design, that is, via anchored items across the test for the different measuring points. In the case of the German test, for t3 the two different tests were linked using the equivalent group design (cf., Kolen & Brennan, 2004). Linking followed the guidelines provided by Kolen and Brennan and was performed via fixed parameters scale linkage within the Rasch model (cf., Bond & Fox, 2007).

Questionnaires with *Likert-scale responses* were identical (in content and wording) at all measuring points. After Rasch-scaling, each scale was calibrated at t2 and then cross-checked at t1 to ensure satisfactory item functioning at both measuring points (they were only used as performance antecedents therefore not measured at t3). If necessary, extreme response categories were collapsed to ensure ordered category thresholds (cf., Bond & Fox, 2007).

Missing data.

Single item non-responses did not constitute a problem because they were imputed by the Rasch procedures when calculating the scale measures (cf., Prieto, Alonso, & Lamarca, 2003) which normally does not make any difference in the final measures obtained (Brentari & Golia, 2008). However, full scale non-completions remained as it was impossible to always assess each measure for each of the 350 students and their parents over all three measuring points. The percentage of missing data due to drop out or illness/absence ranged between 0.3% and 7.1% for t1, reached 2.3% to 11.1% for t2, and 14.9% to 18.9% for t3. For the parents, questionnaire non-response rates were 9.4% for t1 and 22.0% for t2. To account for these full scale non-completions, the dataset underwent multiple imputations (Rubin, 2009; Schafer, 1997) performed in SPSS 18 (IBM, 2009). Five imputed datasets were generated for further analysis, which is estimated to ensure excellent results (Schafer, 1997).

Structural equation models.

The four approaches to giftedness and expertise were modelled as SEMs with the imputed datasets in Mplus 6.11 (B. O. Muthén & Muthén, 1998-2011). As the precondition of multivariate normal distribution of all variables was not met, all models were estimated by MLR, that is, Maximum Likelihood parameter estimation with standard errors that are robust to non-normality (L. K. Muthén & Muthén, 1998-2009; Yuan & Bentler, 2000). As the moderator model included ordinal level variables it was estimated with robust weighted least squares estimators (B. O. Muthén, 1984; B. O. Muthén, du Toit, & Spisic, 1997).

For each model, latent variables were built from the Rasch scaled measures of the scales. The latent factor of school achievement for each measuring point was built from the Rasch measures for English, German and mathematics at the respective measuring point; intelligence was built from the Rasch measures for the verbal, quantitative and non-verbal scale of the KFT4-12+; the factor motivation was built from the Rasch measures of academic self-concept, interest, mastery and performance goal orientation; learning behavior was built from the Rasch measures of goal commitment, effort, concentration, and working attitude; school and classroom climate was built from the Rasch measures for social and learning pressure, student centeredness, and learning climate; familial environment was built from four single items because Rasch measures could not be obtained; and, the Actiotope factor was built from the 10 Rasch measures of the Actiotope scales and the factor score of (latent) achievement as an indicator for the action repertoire which is not covered by the questionnaire scales. All latent factors were tested first in separate confirmatory factor analyses (CFAs) to ensure well-indicated latent constructs prior to testing the structural model. Due to limited space, these results cannot be displayed here but will be provided to the interested reader on request.

First the originally postulated models were tested (see section on approaches to explain extraordinary performance) for both time intervals (t1 → t2, t2 → t3). In cases of bad model fit, alternative models were specified based on the literature. SEMs were compared according to their *BIC* and *AIC,* which are directly comparable across models (lower values indicate better model fit) and their fit statistics. A uniform and thereby comparable method of testing models against each other was not available due to the constraints of imputed data sets, the four different (not hierarchical) models, and models with ordinal data (Levy & Hancock, 2007; L. K. Muthén & Muthén, 1998-2009). The fit statistics used are listed in the following with their critical values for good fit (Hu & Bentler, 1999; Schermelleh-Engel, Moosbrugger, & Müller, 2003):

- χ^2 test: $p > 0.2$ (non-significance indicates that the original and reproduced covariance matrices do not differ; very conservative with large samples)
- *CFI*: > 0.95
- *TLI*: $0.95 < TLI < 1$
- *RMSEA*: < 0.05 to 0.08 (the latter indicates mediocre discrepancies between model and data, Browne & Cudeck, 1993)
- *SRMR*: < 0.08
- *WRMR*: < 0.95 (substitutes SRMR in models with ordinal data, Yu, 2002)

All reported model coefficients are standardized values (using the variances of dependent and independent variables in the model) to make them comparable within a model.

Logistic regression analysis.

After SEMs had explicated which model elements were useful for predictions of future performance, a logistic regression analysis was performed on the most valid predictors taken from all of the four models to see how well high performers at t3 could be identified at t1. High performers were defined as the top 10% of the sample according to the latent achievement measure with the indicators German, English, and mathematics achievement at t3 resulting in $n = 304$ students below the 90^{th} percentile and $n = 33$ above the 90^{th} percentile. $N = 283$ and $n = 31$ offered the data demanded for the analysis. This top 10% of the sample included only high track students (27% of the German student population in the year 2008, Statistisches Bundesamt Deutschland, 2011) therefore representing the top 2.7% of the student population. Additionally our sample comprised more gifted students than a regular high track sample which makes our sample correspond to those students performing at least two standard deviations above the population mean (top 2.3%). Binary logistic regression was performed in SPSS 18 (IBM, 2009) on the dataset including missing values. All predictors were entered into the equation simultaneously. Sensitivity and specificity were both set to 50% to test the predictors' ability to differentiate between high performers and regular performers.

Results

The following sections present the results of testing each of the four models in their postulated form and, if necessary due to bad model fit, also in a modified form. Modifications were derived from the literature as will be outlined. Each model test was performed two-fold, thereby providing a replication. There will always be a model predicting achievement at t2 with the predictors of t1 and one predicting achievement at t3 with the predictors of t2. Finally, the best predictors are evaluated in terms of their ability to identify future high performers.

IQ-Model

The original IQ-model proposed that school achievement can be predicted by general intelligence. As can be seen in Figure 3 a) and b), for both predictions almost all fit indicators were below the critical values (all non-bold-face indices), which means the models did not represent the data appropriately and are therefore not highly valid. This might seem surprising because the regression coefficients of $\beta_{IQ \rightarrow \text{achievement 2}} = 0.89$ and $\beta_{IQ \rightarrow \text{achievement 3}} = 0.86$ indicate high predictive power, which also leads to residuals of the latent achievement factors (broken lined arrow) that are no longer significant. Hence, the variance of the achievement factors is well explained, but the model does not fit the data. This can occur when indicators are strongly correlated but these correlations are not allowed in the model. In this case, the complete covariance is forced into the regression coefficient.

PLACE FIGURE 3 ABOUT HERE

This statistical explanation concurs with criticism of the general intelligence approach, which posits that domain-specific explanations are much more valid than trying to predict specific achievements with a general set of cognitive abilities (Ericsson, Roring, & Nandagopal, 2007; Helmke & Weinert, 1997; Weinert, 1994). The KFT4-12+ takes that into account by using three subscales which are tuned to the domains of language, mathematics and non-verbal logical reasoning. The first can be assumed to correlate with language achievement, that is, English and German measures, while quantitative and non-verbal logical reasoning should be more strongly associated with mathematics achievement. The modified IQ-model allowed all correlations between residuals of the indicators of the predictor and criterion variables to test this assumption. While the latent factors capture the variance common to all indicators, the residuals represent indicator-specific variance, that is, the verbal, quantitative and non-verbal aspects not shared among the three indicators and respectively among the achievement indicators.

Figure 4 a) and b) display the modified IQ-models. Model fit increased which led to good *CFI*, *RMSEA* and *SRMR* for predicting achievement 2 (Figure 4a) and very good fit for predicting achievement 3 where all fit criteria are met (Figure 4b). The comparison with the original models in Figure 3 a) and b) also show that all *AIC*s and *BIC*s are lower (i.e., better) for the modified models. The models still show high regression coefficients between the latent constructs ($\beta_{\text{IQ}\rightarrow\text{achievement 2}} = 0.78$, $\beta_{\text{IQ}\rightarrow\text{achievement 3}} = 0.64$) but a significant residual variance remains in the predicted achievement factor. Some of the residual correlations also attained significance. For the first prediction (Figure 4a), only the residuals of verbal intelligence were significantly correlated with the residuals of the two language measures (English: $r = 0.19$, German: $r = 0.28$). In the prediction for t3 (Figure 4b), the residuals of verbal intelligence were only significantly correlated with the German 3 residuals (r = 0.30), and no longer with English. But the residuals of quantitative intelligence were now correlated with mathematics 3 residuals (r = 0.34), as one would have assumed, while general logical reasoning was not correlated with mathematics achievement. So, after accounting for the general parts of variance in IQ and achievements, we first found English and German correlated residuals in the verbal IQ scale which then disappeared for English – a subject in which there is quick learning progress at the beginning as students are building subject-specific knowledge, and thereby rendering verbal abilities, as measured in the IQ-scale, decreasingly useful. In mathematics, we did not find residual correlations for mathematics 2, probably due to the explanation of mathematics variance through the general prediction (studies often report the highest correlations of IQ-tests with performance in mathematics, Deary, Strand, Smith, & Fernandes, 2007; Rost, 2009). Mathematics 3, however, shows a lower factor loading on the latent achievement factor, indicating that there is more mathematics specific variance which then correlates with the task specific residuals of quantitative reasoning.

PLACE FIGURE 4 ABOUT HERE

Performance Model

The Performance model suggests that achievement measures always contain the results of deliberate practice and therefore can be used as predictors of later achievement measures, as is evident in the expertise approach even though deliberate practice is not directly assessed. Figure 5 displays the results of the original model version for the prediction of achievement 2 (Figure 5a) and achievement 3 (Figure 5b). Both models do not satisfy any of the proposed fit criteria. Additionally, the path coefficients between the latent factors yield unusual values of $\beta_{t1\rightarrow t2}$ = 1.18 and $\beta_{t2\rightarrow t3}$ = 1.21 respectively, leading to negative residual variances ($\zeta_{achievement\ 2} =$ -0.40, $\zeta_{achievement\ 3}$ = -0.46). Additionally, residual variances of the indicators are often very high with levels up to 84 % (German 3). Hence, the models in their general formulation of scholastic achievement are ill-fitted. As for the IQ-models, the reason supposedly lies in indicator specific variance which is forced into the regression of the latent factors. English consistently showed the strongest factor loadings and made up the latent factors to a large extent. Bivariate correlations indicated that English achievement is highly correlated across measuring points with the lowest being $r_{t1\text{-}t3}$ = .75 ($r_{t1\text{-}t2}$ = .80, $r_{t2\text{-}t3}$ = .83) while German and mathematics showed bivariate correlations between .47 and .60 over time. At the same time, the correlations between the three subjects were moderate which should lead to relatively equal factor loadings ($r_{E1\text{-}M1} =$ .31, $r_{E1\text{-}D1}$ = .43, $r_{D1\text{-}M1}$ = .30; $r_{E2\text{-}M2}$ = .41, $r_{E2\text{-}D2}$ = .34, $r_{D2\text{-}M2}$ = .27; $r_{E3\text{-}M3}$ = .38, $r_{E3\text{-}D3}$ = .40, $r_{D3\text{-}M3}$ = .28). Hence, by allowing subject specific residual correlations, the latent factors should become more representative of the achievement in all three subjects and the regression coefficients should decrease.

PLACE FIGURE 5 ABOUT HERE

Following this line of statistical thought and also the literature on the domain specificity of achievement (Ericsson et al., 2007; Helmke & Weinert, 1997; Weinert, 1994), the performance models were recalculated with the residual correlations between the same subjects at different measuring points. As can be seen in Figure 6, both models show satisfying fit indices and the *AIC* and *BIC* improve compared to the original models in Figure 5. However, the *TLI* = 1.002 of the second model (Figure 6b) indicates an over-fit, that is, more relations were specified than were necessary. Despite the fact that all residual correlations turned out to be significant, the regression coefficients are still very high with $\beta_{t1\rightarrow t2}$ = .98 and high with $\beta_{t2\rightarrow t3}$ = 1.01. The factor loadings of English decreased somewhat but remain the highest compared to the other subjects. Even though the latent achievement factors now represent more of a cross-section of the three subjects, the factors remain highly correlated. This shows that the general variance covered by the latent factors depicts quite invariant aspects of achievement. It is unclear which general aspects are captured here, but they are also strongly correlated with IQ as the previous IQ-models showed. So, as expected, prior performance is a very good predictor of future performance, even a collinear one when achievements are subsumed in a latent factor. The domain-specific nature becomes evident in the model improvements but still does not lead to fully-satisfying models with respect to collinearity or the unusual regression coefficients.

PLACE FIGURE 6 ABOUT HERE

Moderator Model

The moderator model assumes an interplay between intelligence and non-cognitive personal as well as environmental factors. As these factors include a large variety of variables, the model is rather complex. Only latent factors which proved valid in the prior CFAs were used to model the SEM so the fit can be interpreted as the fit of the structural model. Although most fit criteria are reasonable, only the *RMSEA* fulfills the demands for good fit for the prediction of achievement 2 (Figure 7). An explanation is provided by the regression coefficients (Figure 7a) which only attained significance for the predictor intelligence while the moderators showed betas around zero (residual correlations between the IQ indicators and achievement indicators were not allowed). Hence, the structure is not at all in line with the theoretical assumptions of an interplay between the precursor abilities (intelligence) and the moderating forces during the learning process. This is further substantiated by the inter-correlations between the predictors (Figure 7b). The model allowed all inter-correlations between the moderators and intelligence assumptions. However, there were no correlations between IQ and motivation, learning behavior, school/classroom climate or familial appreciation of school matters (although all were measured at the same point of time). Obviously, only the moderators (with the exception of the two environmental moderators) are inter-correlated while IQ does not interact with them. This is in line with Gagné's (2013) assumptions of the variable interplay (moderators interact with each other but not with intelligence) or with Perleth's (Heller et al., 2005) assumptions for any developmental stage beyond preschool age where instead of intelligence the mastered performance level should interact with the moderators. Implementing this structure in the model for t2 leads to negligible changes in model coefficients and to slightly better fit indices with $\chi^2(197) = 3553.281$, $p = 0.000$, $CFI = 0.934$, $RMSEA = 0.047$, $WRMR = 1.097$ where only the *RMSEA* meets good-fit criteria.

PLACE FIGURE 7 ABOUT HERE

For the prediction of achievement at t3, the fit criteria turned out similarly with only a good *RMSEA* (cf. Figure 8a; for the model with the better fitting Gagné/Perleth interaction structure fit indices were $\chi^2(197) = 311.297$, $p = 0.000$, $CFI = 0.963$, $TLI = 0.956$, $RMSEA = 0.041$, $WRMR = 0.996$ with *CFI*, *TLI* and *RMSEA* meeting the critical values). The pattern of the regression coefficients and inter-correlations remained almost the same. IQ remained the strongest predictor but with a slightly lower coefficient of $\beta = 0.84$ (compared to $\beta = 0.91$ in the previous model). The moderators still did not reach significance but motivation ($\beta = 0.15$) showed a somewhat higher coefficient compared to the previous model ($\beta = 0.03$). The residual variance of achievement is higher with $\zeta = 0.25$ than it was in the first model with $\zeta = 0.15$. The inter-correlations among the moderator variables showed only minor changes in the values, as can be seen in Figure 8b.

Taken together, the moderator models are dominated by the predictive power of intelligence which declares the variance of the latent achievement factor to a large extent. The moderators do not contribute to the prediction but are correlated among each other while being unrelated to IQ. A huge body of research, which led to the formulation of moderator models in the first place, suggests that the interplay of the moderators is crucial for the current learning process. Our results show that this interplay is widely independent of intellectual abilities.

PLACE FIGURE 8 ABOUT HERE

Systemic Model

As indicated, for the IQ-, performance and moderator models, construct and predictive validity cannot be investigated separately because the predicted variable – achievement – is part of the model itself and therefore cannot be left out to examine construct validity alone. For the systemic model, however, this is possible and was undertaken first. Figure 9 a) and b) display the results of the CFAs for the latent Actiotope factors at t1 and t2. According to the theory (Ziegler, 2005), several variants of modelling the interaction within the Actiotope can be considered although the full interaction between all constructs as postulated cannot be tested. Harder (2012b) found the displayed model to be the best fitting interaction structure based on theoretical considerations with residual correlations allowed between constructs belonging to the same category of Actiotope constituents, namely, the components, developmental mechanisms and system state. The model fit could not be expected to be very good as a number of the specified residual correlations were insignificant but, nevertheless, both models showed satisfactory *RMSEA*s and *SRMR*s (see Figure 9a and b).

For t1, 9 of the 17 residual correlations attained significance while for t2, 4 correlations did so (at the same time more correlations between the categories were found at t2 than at t1, which was tested in a complementary model in Harder, 2012b, allowing for the opposite residual correlations, that is, we observed a shift not a diminution of residual correlations). This shows some interesting facets of the learners' systems. The Actiotope factor already bundles the common variance of all indicating constructs that show positive factor loadings. This corresponds to the intuitive expectation of positive inter-correlations between high goals, high aptitude, a supportive learning environment, high quality feedback, good anticipation, and so on. The residual correlations, however, also point to some side effects not captured in this intuitive understanding. For example, the negative correlation between subjective action space and environment at t1 (r = -.21) could be interpreted as the possibility that a supportive environment takes over the executive functions of the subjective action space so that the student does not need to make her own decisions about learning or does not learn to reflect on her actions.

Another interesting facet of the models is the factor loading of the action repertoire. This indicator was not a self-reported questionnaire measure, as were the other ten, but the latent factor score of the achievement measures. For t1, its loading on the Actiotope factor was clearly non-significant while it turned to $\lambda = 0.11$ with marginal significance in the second model. Obviously some shift of meaning arose in the Actiotope factor, centering it more on achievements, which could be due to the adaptation of the Actiotope to the new learning environment at Gymnasium. The aforementioned shift in the significance of residual correlations also implies an adaptation process of the students.

PLACE FIGURE 9 ABOUT HERE

The systemic model's predictive validity was tested by adding the regression of the latent achievement factor at t2 and t3 on the latent Actiotope factors at t1 and t2 respectively. Both models showed unsatisfactory model fit with $\chi^2(59) = 652.374$, $p = 0.000$, $AIC = 15943$, $BIC = 16174$, $CFI = 0.560$, $TLI = 0.322$, $RMSEA = 0.170$, $SRMR = 0.121$ for the prediction of achievement at t2 and $\chi^2(59) = 309.369$, $p = 0.000$, $AIC = 16233$, $BIC = 16464$, $CFI = 0.676$, $TLI = 0.500$, $RMSEA = 0.110$, $SRMR = 0.107$ for the prediction of achievement at t3. This is due to the rather low fit of the latent Actiotope factor model, as outlined previously, but also the regression coefficients were non-significant, which disqualified the prediction model. Interestingly, however, the regression coefficient for the prediction of achievement 2 was $\beta = 0.00$ (leaving 100% of the latent factor's variance unexplained) while it attained marginal significance in the model for t3 with $\beta = 0.18$ (explaining 3% of the latent factor's variance), again supporting the idea of a more achievement-centered Actiotope after one year in the new school environment. In the same vein, the factor loading of the action repertoire at t2 (predicting achievement at t3) attained significance in the prediction model compared to the model in Figure 9b with $\lambda = 0.14$. The rest of the model showed no noteworthy changes to the CFAs displayed in Figure 9.

Prediction of Future High-Performance

According to the SEM analyses only intelligence and achievement qualified as good predictors of future performance. Due to the collinearity problems of achievement predictions with the latent achievement factors, German, English, and mathematics performance were entered into the binary logistic regression analysis as manifest variables, as was IQ (high performers were defined according to the latent factor scores at t3). Standard errors of the predictors fell below the value of 1 indicating that multicollinearity was resolved by this means. The regression yielded a highly significant Nagelkerke $R^2 = .66$ and $\chi^2(4) = 118.46$, $p < .001$ classifying 93.6% of the students correctly: 97.2% of the lower achieving students were identified correctly and 61.3% of the high performers. Relevant predictors at t1 were only English (*OR* = 44.52, $p < .001$) and mathematics achievement (*OR* = 4.33, $p < .001$) while German achievement (*OR* = 1.85, $p = .13$) and IQ (*OR* = 0.56, $p = .38$) did not contribute significantly to the classification. The high *OR* of English can be explained by the higher factor loadings of English on the latent achievement factor at t3 ($\lambda = .73$) compared to the loadings of mathematics and German ($\lambda = .47$ and $\lambda = .43$ respectively).

Discussion

This study aimed to empirically examine and compare four approaches to giftedness and expertise, which were operationalized as the IQ-model, the performance model, the moderator model, and the systemic model. Each approach was modelled with data from a longitudinal study and replicated for two different intervals between measuring points trying to fulfill the standards for a theory comparison set by the literature. The following sections will first discuss the results obtained for each model, addressing the first research question of the construct and predictive validity of each model and, thereby, each theoretical approach. In this context, the second research question of change over time and over students' development for each model will also be discussed. Finally, the approaches will be compared to answer the research question of which approach is best suited for which purposes.

Validity of the Four Approaches and Change over Time

Validity of the IQ-approach.

The original IQ-model, which assumed that general intelligence predicted achievement, did not prove valid as the models did not fit the data well. However, the models improved dramatically when domain-specificity was taken into account by allowing correlations between the IQ-subscales and the separate achievement measures. These adapted models showed IQ to be a very good predictor of achievement but also had different domain-specific developments in the prediction of achievement at t2 as compared to t3. Domain-specific residual correlations shifted from verbal-IQ with German and English (students started learning English intensively with the beginning of secondary education, i.e., t1) for the prediction of achievement 2 to correlations of verbal-IQ with German and quantitative-IQ with mathematics for the longer time-interval prediction. English residuals were no longer correlated with the more general language abilities measured by the verbal IQ scale, which makes sense as students progress quickly and English knowledge becomes increasingly specific. Mathematics residuals became correlated with quantitative reasoning residuals but remained uncorrelated with general logical reasoning residuals, which can be explained by the more challenging and specific math problems that students meet in class according to the curriculum (Institut für Schulqualtiät und Bildungsforschung München, 2004). Given this trend of specialization of English abilities and mathematics challenges moving away from the more general IQ-measures within only two years of schooling, one can expect much heavier specializations in the upcoming years (Ericsson et al., 1993; Sternberg, 2001).

With regard to the *construct validity* of the IQ-approach, we conclude that the original version with the assumption of general intelligence as a precursor of later achievement does not withstand the empirical test. A more domain-specific understanding of intelligence, however, can solve this problem. The second consideration concerns the observed dynamic in domain-specific relationships between intelligence sub-abilities and specific academic domains. Obviously, intelligence-associated abilities cannot be perceived of as stable correlates of achievement. This raises the question whether it then makes sense to assume intelligence-associated abilities as precursors, such that when they have to be domain-specific and dynamic, one can directly use performance measures. This supports research on expert performance, which points out that, with growing expertise, general abilities become less conducive to competent action while being replaced with specific knowledge and routines of the domain to assure efficient handling of domain-specific problems (Ericsson et al., 2007; Helmke & Weinert, 1997).

The *predictive validity* of general intelligence, on the other hand, was significant for both time intervals ($\beta = 0.78$ and $\beta = 0.64$) although losing some predictive power (probably) to the intensifying domain-specificity over the two years. Nevertheless, intelligence remains a rather good predictor which is supported by manifold evidence, for example, for predictions of school performance a few years later (Deary et al., 2007; Leeson, Ciarrochi, & Heaven, 2008; Renzulli, 2005) or study performance (Hell, Trappmann, & Schuler, 2007; Kuncel, Hezlett, & Ones, 2004) while it is normally less predictive for professional success (Ng, Eby, Sorensen, & Feldman, 2005; Rost, 2009; Strenze, 2007).

Validity of the performance approach.

The SEMs of the performance model were very similar for both measuring points. Both struggled with collinearity of the latent factors but at the same time this collinearity buttresses the correlation between achievements at various points in time and thereby their usefulness as predictors. The correlations between achievements of the same kind also became apparent in the domain-specific bivariate correlations over time which supports extant literature (Helmke & Weinert, 1997; Ree, Caretta, & Teachout, 1995). Again, the general formulation of achievement (as before of intelligence) did not prove valid. Domain-specific residual correlations improved the models (although still struggling with collinearity) and were all significant, again pleading for specific rather than general approaches, as proposed by Ericsson and colleagues (2007), for example.

With regard to the *construct validity* of the performance approach, these results do not provide a definitive answer. First, the generality problem of the achievement modelling suggests a much more specific investigation of the performance approach to expertise. Second, the study limitations did not allow us to examine the causal role of deliberate practice explicitly as would be demanded by a thorough test of the approach. Third, the limitations to the development of expertise discussed by Ericsson, Krampe and Tesch-Römer (1993) were not investigated either.

The *predictive validity* of the approach, however, was well supported. Prior achievement turned out to be the only valid predictor of future high performers, resulting in at least 61% high performers being correctly identified. Achievement was, therefore, the best measure for predictions over one year, irrespective of which measuring points in this study were examined. The predictive power of prior performance has been widely recognized in theories of giftedness and expertise development and is strongly supported by studies from many contexts, for example, in higher education (Cassidy, 2012; Richardson, Abraham, & Bond, 2012).

Validity of the moderator approach.

The moderator models for both measuring points were very similar. Intelligence played the dominant role in predicting future achievement while the moderators, which were meant to be the innovative add-on of the approach, did not predict achievement. However, they correlated quite well among each other, which points towards the interplay of motivation, learning behavior, school/class climate and the appreciation of school matters at home in the current learning process. Notably, though, none of them correlated with IQ at all. Heller (2005) assumes that precursor gifts such as intelligence interact with the moderators during the learning process, which was not substantiated by our results. While intelligence is the starting point of development that could follow many different paths, the moderators at play during the learning process constitute a snapshot of the current situation and are therefore likely to change according to the prevalent dynamics. This makes a correlation with intelligence unlikely and is also considered in several other theories. According to Ericsson and colleagues (2007) the current learning process, or development, is based on specific acquired knowledge and abilities, not some general cognitive abilities. In his moderator conception, Perleth (Heller et al., 2005) takes this into account by regressing learning processes after the initial steps in a domain (i.e., from primary school on) on acquired domain-specific precursor abilities and knowledge for the next developmental steps, instead of always regressing on general precursors such as intelligence. It is possible that, in very early learning processes, intelligence plays the suggested role but even for kindergarten children intelligence as a predictor of math and language performance is superseded by other predictors, especially by measures of executive functioning (Roebers, Röthlisberger, Cimeli, Michel, & Neuenschwander, 2011; Roebers et al., 2014). In the same vein, Eckstein (2000) showed that the influence of intelligence diminishes dramatically to a non-significant level as soon as an ability is trained, which actually happens very early in childhood. So, Gagné's (2013) formulation might be the most valid, although it is inaccurate on the role played by intelligence. He assumes that the moderators influence the learning process but do not correlate with intelligence or other general abilities irrespective of which developmental stage is considered.

The model tests also encountered a methodological problem worth discussing. The influence of moderator variables on achievement has been shown in several studies (Campbell & Kyriakides, 2011; Cho, Lin, & Hwang, 2011; Heller, 2001; Helmke & Weinert, 1997; Simonton, 1994). The fact that no moderator showed significant regression coefficients in the present study might be due to a problem of operationalization. IQ-tests are designed to predict achievement and hence show high predictive power. Moderators, on the other hand, were assessed by self-report questionnaires and are less related to achievement. More objective, reliable and valid measures of moderator variables would be necessary to even out the imbalance in potential predictive power.

With regard to the *construct validity* of the moderator approach, we conclude that it is not attained with intelligence as the precursor gift at the developmental stage that we investigated. It is also not attained at earlier stages because training of abilities begins very early and dramatically reduces the influence of intelligence on current learning dynamics. The moderators, on the other hand, can be viewed as a promising theoretical add-on of the moderator approach as the model fits improved, especially for the prediction of achievement

3. Here the moderators gained weight in the model, reducing the influence of IQ, which might be associated with a better established interplay of non-cognitive and environmental factors with the students' increasing adaptation to the new school and learning situation.

The *predictive validity* of the moderator models cannot be attributed to the moderator approach but is simply due to the predictive power of intelligence. The clue of the moderator models – the moderator variables – does not improve the predictions of future achievement. Although the moderator of motivation gains some influence for the prediction of achievement 3, downgrading the influence of IQ, it still does not contribute significantly to the prediction. This is unsatisfactory in that it unveils a basic conceptual problem of the moderator approach.

Validity of the systemic approach.

The models adhering to the systemic approach revealed that many interactions exist between the constituting variables of the system as defined by the Actiotope theory (Ziegler, 2005). The residual correlations provided some interesting insights. Given that the common aspects of all Actiotope entities working in the same direction were already summarized in the latent factor, the residual correlations pointed to more complex and sometimes even counterintuitive side effects, which should be studied in more detail. Interestingly, those correlations shifted from the first to the second model. The interplay of Actiotope components, developmental mechanisms and system state therefore build a very dynamic constellation. The shift of inter-correlations probably represents the adaptation of the students to their new school environment, which might also be the source of the higher centering of the Actiotope on performance (evident in the higher factor loading of prior performance on the Actiotope factor at t2). The stronger focus on performance in the Actiotope also came with not yet significant but higher predictive power of the Actiotope for future performance.

The modelling of the Actiotope also posed fundamental methodological problems, which led to mediocre model fit. The first problem consisted in the modelling of the full interaction among all constructs, which was partly solved by including one part of the residual correlations (those among constructs of the same category, i.e., components, developmental mechanisms, system state). Second, the theory proposes non-linear relationships among the Actiotope entities that could not be modeled in the SEMs. Third, it assumes highly individual system constellations and developmental trajectories, which questions the results of a group analysis. So, different means of analysis are needed to propel research on systemic models such as this. Another methodological problem concerned assessment methods. The differently-assessed action repertoire (test data) did not correspond well with the self-report questionnaire data (cf. IQ vs. moderator assessment discussion in the previous paragraph).

Regarding *construct validity*, the systemic approach showed some promising modelling of variable interplay emphasizing the dynamic nature of a learner's system and the challenges to its adaptability. However, the methodological problems render a thorough estimation of its validity very difficult and which should be addressed in future studies.

The *predictive validity* of the systemic models was low as the regression coefficients at t1 were not significant and at t2 reached marginal significance explaining 3% of the achievement variance. Considering that the Actiotope variables capture a state of a very dynamic interplay and relate this snapshot to achievement one year later, this is not overly surprising. Nevertheless, this is an unsatisfactory prediction but with a tendency towards improvement. The Actiotope has successfully been used to predict school achievement and other student variables in grades 7 and 8 in three different countries (Ziegler et al., 2014), so the predictive power probably increases with advancement in the school trajectory and the accompanying adaptation of the Actiotope to the demands of the school environment.

Comparison of the Four Approaches

The previous discussion already indicated that none of the approaches in their current form or their methodological possibilities of investigation are perfectly valid. Hence a comparison of the approaches needs to consider strengths and weaknesses in relation to certain goals that the approaches are supposed to serve (Hacking, 1983; Suppe, 1977). The goals of the theories of giftedness and expertise have already been introduced as explaining and predicting future achievement. The first of these searches for valid explanations, specifying factors of influence and their interactions during the process of developing expertise (Davidson & Downing, 2000; Helmke & Weinert, 1997) thereby providing possibilities for interventions. The latter goal, prediction, is more oriented towards simple diagnostic needs of selecting students and the like. Sometimes a valid predictor is sufficient to select a group of students for a program, for example, without knowing why this predictor leads to successful participation in the program or which other factors might be involved. These are simpler tasks, in practice, while the more complex ones, such as fostering a student over several years and handling problems that occur along the way, also affords a thorough explanation of the developmental process to be able to steer it (cf. for example current counseling practice, Harder, 2012a; Ziegler, Grassinger, Stoeger, & Harder, 2012).

Suitability of the approaches for predicting school performance.

We will examine in more detail the easier challenge for the theoretical approaches, namely that of predicting future performance. To make good predictions, a theory only has to fulfill the demands of predictive validity. Of the four approaches put to the test, only the IQ- and the performance approaches allowed for good performance predictions. Compared to the moderator and systemic approaches, this could be expected as the latter captured a snapshot of a dynamic interplay, which is not as predictive as a more stable variable like IQ or performance; furthermore, the moderator and systemic models had to deal with less powerful measuring instruments of a greater number of variables. Nevertheless, the IQ-model and performance model differed greatly in their predictions, clearly favoring the performance measures as predictors of future performances. Most likely, this draws back on the higher domain specificity of the latter approach which repeatedly proved to be crucial. Ericsson and colleagues (2007) state that there is no better predictor of future performance than current performance, which seems quite logical given that many of the competences necessary for future success are integrated in the current achievement. However, others (Neubauer & Opriessnig, in press; Rost, 2009) argue that intelligence is the strongest predictor of future success. Lohman's (2005) finding that a distal measure like IQ is better suited to predict long-term success, while proximal measures like current performance work better in the short term, suggests that these positions can be reconciled. For our measuring points lying within three school years, the time interval favors the proximal performance measures as predictors. When predicting final exam grades or study success, IQ might turn out to provide the more valid predictions, as long as the criterion does not depend on strong specialization in a domain like vocational success (cf. the discussion of the IQ-approach). The decision for a predictor therefore depends on the criterion one wants to predict and the time interval one wants to cover. Practitioners in education predominantly need to match learners with special instruction. In these cases, we need to predict the proximal success of a student facing the demands of a (hopefully) clearly defined provision, so prior performance in tasks with similar demands would be the suggested predictor and criterion for accepting students to that provision.

This line of thought was corroborated by the overall analysis of available predictors we carried out – the *identification of future high performers*. We aimed at predicting high performance in the main subjects, thus general academic abilities rather than special abilities were in focus, which might be interesting for granting scholarships or access to interdisciplinary provisions. At t3, 61.3% of the high performers could be identified by using performance and IQ measures from t1. As could be expected, only performance measures (English and mathematics achievement) contributed to this prediction whereas IQ did not. This is the first noteworthy finding in this analysis because intelligence is the most widely used identification measure (Ziegler, 2008b). The lack of ability to differentiate individuals at high performance levels is a known fact about IQ-measures (Ericsson, Nandagopal, & Roring, 2008; Feldman, 1984; Gruber & Ziegler, 1996; Schneider, 2002; Subotnik, Kassan, Summers, & Wasser, 1993), but practitioners in search for alternatives stick to the well-regarded IQ-tests. The second noteworthy finding concerns the ratio of non-identified high performers. Despite using the best predictors available (which were quite substantial considering the construct coverage in our study), about 40% of the target students for special provisions remained undetected. This represents a rather unacceptable result (Schendera, 2008), which means that even better predictors need to be found. Several studies have yielded good predictive results with motivational variables (Dresel, Fasching, Steuer, & Berner, 2010; Gottfried, Gottfried, Cook, & Morris, 2005) or the socioeconomic status of the parents (Holohan & Sears, 1995; Strenze, 2007), which might statistically improve such identification procedures. However, in the case of SES, such modification is highly questionable from an ethical point of view. A new and holistic approach uses teacher ratings of students' resources for learning (learning and educational capital, Ziegler & Baker, 2013) to determine the needs that are critical for fostering the advancement of their performance levels. Teacher ratings of learning resources, in one study, correlated very well with objective performance measures and school grades (Harder, Trottler, & Ziegler, 2013); notably, the correlations of up to r = 72 with an objective performance test were comparable to the predictions of IQ-tests. Although their incremental value above prior performance measures in a longitudinal prediction remains to be demonstrated, this seems to be a promising approach.

Suitability of the approaches for explaining school performance.

The second purpose of theories is to explain the phenomenon at hand, in our case the development of expertise. As has been outlined previously, to serve as a good explanation a theory has to fulfill content, construct and predictive validity. *Content validity* was not examined in detail in this study but can be summed up in the following (cf. Harder, 2012c). The IQ-approach was the historical precursor of the other approaches and had serious extensions added to it to ensure higher content validity, consequently the IQ-approach has low content validity. The performance approach only offers some vague limiting factors to the development of expertise, which lowers its content validity. The moderator approach and the systemic approach, however, include all conceivable factors of influence on the developmental process and therefore can be viewed as highly content valid. Concerning the second demand of *construct validity*, the IQ- and performance models also failed due to their general instead of domain-specific formulation. Applied to a specific domain instead of general school performance, the performance model might have fulfilled construct validity but the factors of the theory left out in our analysis would still have to prove valid. The moderator approach struggled with the conceptual problem of assuming intelligence to interact with the moderating forces during the advanced learning processes in secondary education, which rendered it construct invalid. The systemic approach fulfilled construct validity to the extent that it could be modeled. In this case, methodological constraints prevented the full test of construct validity. The third demand, *predictive validity*, has been discussed above, concluding that only the IQ- and performance approaches can be deemed predictively valid.

In conclusion, currently no approach fulfills all three demands to serve as a good explanation of expertise development. The IQ- and performance approaches were neither content nor construct valid but predictively valid, whereas the moderator approach was only content valid, and the systemic approach was content and construct valid but not predictively valid. In their current iterations, the systemic approach is the most promising candidate for a good explanation but still needs to overcome methodological problems in assessment and modeling. It also offers the soundest theoretical basis as was outlined in the introduction of the historical development of the approaches. Interestingly, the moderator approach and performance approach could also reach construct validity by a conceptual revision which would make them approach each other: the performance approach lacks a clear depiction of moderating factors of influence while the moderator approach needs more proximal precursor abilities and clearer interaction structures between moderators (and powerful assessment possibilities). With a further developed systemic approach and a revised, maybe even combined moderator-performance approach, content and construct validity could be assured. Moreover, predictions should become possible with more suitable modeling options (especially for systemic assumptions) and on shorter time intervals where the modeled dynamics of the learning processes are still closely linked to their outcomes. Such conceptions of giftedness and expertise would enhance our understanding of the learning processes necessary to build up excellent knowledge and abilities in a domain and hence would empower practitioners to assure highly effective interventions and instruction according to each student's individual constellation of influencing factors.

Limitations

Besides the limitations concerning single models that have already been mentioned in the respective discussions, the present study naturally suffered from some general limitations. From a methodological perspective, the first limitation entails the treatment of the nested data structure. Hierarchical analyses were indicated according to the obtained intra-class correlations and should be considered as listed in the criteria for an empirical theory comparison. However, it was impossible to consider them due to the limited sample size of 16 clusters — that is, classrooms — which should be 30–50 (Division of Statistics and Scientific Computation - University of Texas at Austin, n.d.). Furthermore the number of clusters limits the number of parameters that can be estimated to 15, which was greatly exceeded by the moderator- and systemic models. Second, the chosen method of building general achievement factors revealed some problems. The latent factors covered rather little of the indicators' variance, which resulted in high residuals. Accordingly, the latent factors were highly correlated with each other, even multicollinear. Therefore, it would be better to resign from the latent factor modeling of general achievement. For example, Eid (2000) and Geiser (2010) suggest the modeling of indicator (in our case domain) specific residual factors as a solution for cases in which one indicator can be prioritized and treated as a reference indicator. Unfortunately, this could not be legitimized theoretically in the present study.

A third limitation concerns the interpretational scope of the study, which is limited to the four approaches tested with some further limitation due to the incomplete testing of the performance approach. As the multi-component approach was left out completely, no judgment can be drawn on its usability for scientific or practical purposes.

Future studies should consider these problems when planning an empirical comparison. Putting competing conceptions up against each other can provide very useful information, far surpassing the conclusions drawn from single evaluation studies. To our knowledge, this was the first study drawing such a comparison to evaluate different theories' validity and, in so doing, has led to many interesting findings and points to stimulate further developments of conceptual aspects, methodological possibilities and practical implications.

References

Bäuerlein, K., Lenhard, W., & Schneider, W. (2012). *LESEN 8-9, Lesetestbatterie für die Klassenstufen 8-9* [Reading test battery for grades 8 to 9]. Göttingen: Hogrefe.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by reseachers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*(2), 161-186.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). New York: Routledge.

Brentari, E., & Golia, S. (2008). Measuring job satisfaction in the social services sector with the Rasch model. *Journal of applied measurement*, *9*(1), 45-56.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park: Sage.

Campbell, J. R., & Kyriakides, M. L. (2011). Applying the Munich Dynamic Ability-Achievement Model. In A. Ziegler & C. Perleth (Eds.), *Excellence* (pp. 283-302). Berlin: Lit.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Cassidy, S. (2012). Exploring individual differences as determining factors in student academic achievement in higher education. *Studies in Higher Education, 37*(7), 793-810. doi: 10.1080/03075079.2010.545948

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of educational psychology, 54*, 1-22.

Cho, S., Lin, C.-Y., & Hwang, D. (2011). Predictive relationships among variables on math talent development based on the Munich Dynamic Ability-Achievement Model. In A. Ziegler & C. Perleth (Eds.), *Excellence* (pp. 268-282). Berlin: Lit.

Csikszentmihalyi, M., & Wolfe, R. (2002). New conceptions and research approaches to creativity: Implications of a systems perspective for creativity in education. In K. A. Heller, F. J. Mönks, R. J. Sternberg & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (pp. 81-94). Oxford: Elsevier.

Cury, F., Elliot, A. J., Da Fonseca, D., & Moller, A. C. (2006). The social-cognitive model of achievement motivation and the 2 x 2 achievement goal framework. *Journal of Personality & Social Psychology, 90*(4), 666-679.

Dai, D. Y., & Renzulli, J. S. (2008). Snowflakes, living systems, and the mystery of giftedness. *Gifted Child Quarterly, 52*(2), 114.

Davidson, J. E., & Downing, C. L. (2000). Contemporary models of intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 34-52). New York, NY: Cambridge University Press.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13-21.

Division of Statistics and Scientific Computation - University of Texas at Austin. (n.d.). Mplus for Windows: An introduction. Retrieved July 14th, 2011, from http://ssc.utexas.edu/images/stories/ssc/files/tutorials/MPlus_Tutorial.pdf

Dresel, M., Fasching, M., Steuer, G., & Berner, V. D. (2010). The role and the interplay of classroom goal structures, individual motivation and general intelligence for (excellent) school achievement in elementary school classrooms. *Talent Development & Excellence, 2*(1), 63-81.

Eckstein, S. G. (2000). Growth of cognitive abilities: Dynamic models and scaling. *Developmental Review, 20*, 1-28.

Eder, F., & Mayr, J. (2000). *Linzer Fragebogen zum Schul- und Klassenklima für die 4. bis 8. Klassenstufe (LFSK 4-8)* [Linz questionnaire on school and classroom climate for grades 4 to 8]. Göttingen: Hogrefe.

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika, 65*, 241-261.

Elliot, A. J. (1997). Integrating the classic and contemporary approaches to achievement motivation: A hierarchical model of approach and avoidance achievement motivation. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 143-179). Greenwich, CT: JAI Press.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychologial Review, 100*(3), 363-406.

Ericsson, K. A., Nandagopal, K., & Roring, R. W. (2008). An expert-performance approach to the study of giftedness. In L. V. Shavinina (Ed.), *Handbook on giftedness*. New York, NY: Springer.

Ericsson, K. A., Roring, R. W., & Nandagopal, K. (2007). Giftedness and evidence for reproducibly superior performance: An account based on the expert performance framework. *High Ability Studies, 18*(1), 3-56.

Feldman, D. H. (1984). A follow-up of subjects scoring above 180 IQ in Terman's 'Genetic Studies of Genius'. *Exceptional Children, 50,* 518-523.

Gagné, F. (1985). Giftedness and talent: Reexamining a reexamination of the definitions. *Gifted Child Quarterly, 29,* 103-112.

Gagné, F. (2005). From gifts to talents: The DMGT as a developmental model. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 98–119). New York, NY: Cambridge University Press.

Gagné, F. (2013). The DMGT: Changes within, beneath, and beyond. *Talent Development & Excellence, 5*(1), 5-19.

Geiser, C. (2010). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung* [Applied data analysis with Mplus. An introduction]. Wiesbaden: Springer.

Gottfried, A. W., Gottfried, A. E., Cook, C. R., & Morris, P. E. (2005). Educational characteristics of adolescents with gifted academic intrinsic motivation: A longitudinal investigation from school entry through early adulthood. *Gifted Child Quarterly, 49,* 172-186.

Gruber, H., & Ziegler, A. (Eds.). (1996). *Expertiseforschung. Theoretische und methodische Grundlagen* [Research on expertise. Theoretical and methodological basics]. Opladen: Westdeutscher Verlag.

Hacking, I. (1983). *Representing and intervening. Introductory topics in the philosophy of natural science*. Cambridge, UK: Cambridge University Press.

Harder, B. (2012a). The current need for a system-theoretical background in counselling the gifted. *High Ability Studies, 23*(1), 69-71.

Harder, B. (2012b). An empirical validation of the Actiotope Model of Giftedness. In H. Stoeger, A. Aljughaiman & B. Harder (Eds.), *Talent Development and Excellence* (pp. 31-42). Münster: LIT.

Harder, B. (2012c). *Modelle zur Erklärung von Leistungsexzellenz im theoretischen und empirischen Vergleich* [Conceptions to explain achievement excellence – a theoretical and empirical comparison]. Münster: Lit.

Harder, B., Trottler, S., & Ziegler, A. (2013). Die Entwicklung der Nürnberger Bildungs- und Lernkapital Checkliste (NBLC) [The development of the Nuremberg Checklist for Educational and Learning Capital]. *Journal für Begabtenförderung, 13*(2), 37-47.

Harder, B., & Ziegler, A. (2009). *Englisch Leistungstest für die 5. bis 7. Jahrgangsstufe* [English achievement test for 5th to 7th grade]. Unpublished test battery. University of Ulm. Ulm, Germany.

Hell, B., Trappmann, S., & Schuler, H. (2007). Eine Metaanalyse der Validität von fachspezifischen Studienfähigkeitstests im deutschsprachigen Raum [A meta-analysis of the validity of domain-specific study-capacity-tests in German-speaking areas]. *Empirische Pädagogik, 21*, 251-270.

Heller, K. A. (Ed.). (2001). *Hochbegabung im Kindes- und Jugendalter* [Giftedness in childhood and adolescence] (2nd ed.). Göttingen: Hogrefe.

Heller, K. A., Mönks, F. J., Sternberg, R. J., & Subotnik, R. F. (Eds.). (2002). *International handbook of giftedness and talent* (2nd ed.). New York, NY: Elsevier.

Heller, K. A., & Perleth, C. (2000). *KFT 4-12+ R, Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision* [Cognitive abilities test for 4th to 12th grad, revision]. Weinheim: Beltz.

Heller, K. A., Perleth, C., & Lim, T. K. (2005). The Munich model of giftedness designed to identify and promote gifted students. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 147-170). New York, NY: Cambridge University Press.

Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen [Determinants of school performance]. In F. E. Weinert (Ed.), *Enzyklopädie der Psychologie, Serie Pädagogische Psychologie, Vol. 3: Psychologie des Unterrichts und der Schule* [Encyclopedia of psychology, Vol. 3: Psychology of instruction and school] (pp. 71-176). Göttingen: Hogrefe.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research, 60*, 549-571.

Holohan, C. K., & Sears, R. R. (1995). *The gifted group in later maturity*. Stanford, CA: Stanford University Press.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of educational psychology, 57*, 253-270.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance struture analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

IBM. (2009). SPSS Statistics 18, Version 18.0.0. [Computer program]. Ehningen: IBM Deutschland GmbH.

Institut für Schulqualtiät und Bildungsforschung München, I. S. B. (2004). Das Gymnasium in Bayern [The Gymnasium in Bavaria (currucula)]. Retrieved July 12th, 2014, from http://www.isb-gym8-lehrplan.de/contentserv/3.1.neu/g8.de/data/media/26418/Lehrplaene/Gym_in_Bay_Fachprofile.pdf

Kaplan, D. (2000). *Structural equation modeling. Foundations and extensions*. Thousand Oaks, CA: Sage.

Keller, G., & Thiel, R.-D. (1998). *Lern- und Arbeitsverhaltensinventar (LAVI)* [Inventory on learning and working behavior]. Göttingen: Hogrefe.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*(1), 148-161.

Leeson, P., Ciarrochi, P. L., & Heaven, P. C. L. (2008). Cognitive ability, personality, and academic performance in adolescence. *Personality and Individual Differences, 45,* 630-635.

Levy, R., & Hancock, G. R. (2007). A framework of statistical tests for comparing mean and covariance structure models. *Multivariate Behavioral Research, 42*(1), 33-66.

Linacre, J. M. (2011). WINSTEPS® (Version 3.71.0.1). [Computer program] Beaverton, OR: Winsteps.com.

Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly, 49*(2), 111-138.

Marsh, H. W. (1990). *Self Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept: An interim test manual and a research monograph*. San Antonio, TX: Psychological Corporation.

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

Mönks, F. J., & Katzko, M. W. (2005). Giftedness and gifted education. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 187-200). New York, NY: Cambridge University Press.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115-132.

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript.

Muthén, B. O., & Muthén, L. K. (1998-2011). Mplus Version 6.11. Base program and mulitlevel add-on. [Computer program]. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998-2009). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Neubauer, A. C., & Opriessnig, S. (in press). The development of talent and excellence – Do not dismiss psychometric intelligence, the (potentially) most powerful predictor. *Talent Development & Excellence, 6*(2).

Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology, 58*, 367-408.

Pekrun, R., Götz, J. S., Zirngibl, A., vom Hofe, R., & Blum, W. (2002). *Skalenhandbuch PALMA 1. Messzeitpunkt (5. Klassenstufe)* [Handbook of scales used in PALMA measuring point 1 (fifth grade)]. München: Universität München, Institut Pädagogische Psychologie.

Preckel, F. (2008). *Skala familiäre Wertschätzung* [Scale of familial appreciation]. Unpublished questionnaire, Universität Trier. Trier, Germany.

Prieto, L., Alonso, J., & Lamarca, R. (2003). Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health and Quality of Life Outcomes, 1,* 1-27.

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests*. Copenhagen: Danish Institute for Educational Research (Expanded edition, 1980. Chicago: University of Chicago Press).

Ree, M. J., Caretta, T. R., & Teachout, M. S. (1995). Role of ability and prior knowledge in complex training performance. *Journal of Applied Psychology, 80*, 721-730.

Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinisc and extrinsic motivation: The search for optimal motivation and performance* (pp. 373-404). San Diego, CA: Academic Press.

Renzulli, J. S. (1986). *Systems and models for developing programs for the gifted and talented*. Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S. (2005). The three-ring conception of giftedness: A developmental model for promoting creative productivity. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 246-280). New York, NY: Cambridge University Press.

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

Roebers, C. M., Röthlisberger, M., Cimeli, P., Michel, E., & Neuenschwander, R. (2011). School enrolment and executive functioning: A longitudinal perspective on developmental changes, the influence of learning context, and the prediction of pre-academic skills. *European Journal of Developmental Psychology, 8*(5), 526-540.

Roebers, C. M., Röthlisberger, M., Neuenschwander, R., Cimeli, P., Michel, E., & Jäger, K. (2014). The relation between cognitive and motor performance and their relevance for children’s transition to school: A latent variable approach. *Human movement science, 33*, 284-297.

Rost, D. H. (2009). *Intelligenz: Mythen und Fakten* [Intelligence: myths and facts]. Weinheim: Beltz/Psychologie Verlagsunion.

Rubin, D. B. (2009). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Schendera, C. (2008). *Regressionsanalyse mit SPSS* [Regression analysis with SPSS]. München: Oldenbourg.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research - Online, 8*(23-74).

Schneider, W. (2002). Giftedness, expertise, and (exceptional) performance: A developmental perspective. In K. A. Heller, F. J. Mönks, R. J. Sternberg & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (2nd ed.). New York, NY: Elsevier.

Simonton, D. K. (1994). *Greatness: Who makes history and why*. New York, NY: Guilford.

Souvignier, E., Trenk-Hinterberger, I., Adam-Schwebe, S., & Gold, A. (2008). *FLVT 5-6: Frankfurter Leseverständnistest für 5. und 6. Klassen* [Frankfurt reading comprehension test for 5th to 6th grade]. Göttingen: Hogrefe.

Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology, 15*, 201-293.

Statistisches Bundesamt Deutschland. (2011). Allgemeinbildende Schulen - Schüler/innen nach Schularten [School education - students by sort of school (census data)]. Retrieved August 2nd, 2011, from http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Statistiken/BildungForschungKultur/Schulen/Tabellen/Content75/AllgemeinbildendeSchulenSchulartenSchueler,templateId=renderPrint.psml

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.

Sternberg, R. J. (2001). Giftedness as Developing Expertise: a theory of the interface between high abilities and achieved excellence. *High Ability Studies, 12*(2), 159-179. doi: 10.1080/13598130120084311

Sternberg, R. J. (2005). The WICS model of giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 327-342). New York, NY: Cambridge University Press.

Sternberg, R. J., & Davidson, J. E. (Eds.). (2005). *Conceptions of giftedness* (2nd ed.). New York, NY: Cambridge University Press.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35*(5), 401-426.

Stumpf, E. (2008). *Fragebogen zum selbstregulierten Lernen* [Questionnaire on self-regulated learning]. Unpublished questionnaire, Julius-Maximilians-Universität Würzburg. Würzburg, Germany.

Subotnik, R. F., Kassan, L., Summers, E., & Wasser, A. (1993). *Genius revisited: High IQ children grown up*. Norwood, NJ: Ablex.

Suppe, F. (1977). *The structure of scientific theories*. Urbana, IL: University of Illinois Press.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.

Weinert, F. E. (1994). Lernen lernen und das eigene Lernen verstehen [Learning to learn and understanding one's learning]. In K. Reusser & M. Reusser-Weyeneth (Eds.), *Verstehen. Psychologischer Prozess und didaktische Aufgabe* [Understanding. Psychological process and didactic challenge] (pp. 183-205). Bern: Huber.

Weiß, C., & Schneider, W. (2009). *Mathematiktestinventar für 5. bis 7. Jahrgangsstufe* [Mathematics achievement test inventory for 5th to 7th grade]. Unpublished test battery. Julius-Maximilians-Universität. Würzburg. Würzburg, Germany.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68-81.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes.* Doctoral dissertation. University of California, Los Angeles.

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In M. E. Sobel & B. M. P (Eds.), *Sociological methodology 2000* (pp. 165-200). Wahington, D.C.: ASA.

Ziegler, A. (2005). The Actiotope model of giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 411-436). New York, NY: Cambridge University Press.

Ziegler, A. (2008a). *Fragebogen zur Erfassung des schulischen Aktiotops* [Questionnaire to assess the school-related Actiotope]. Universität Ulm: Unpublished questionnaire. Ulm, Germany.

Ziegler, A. (2008b). *Hochbegabung* [Giftedness]. München: Reinhardt.

Ziegler, A., & Baker, J. (2013). Talent development as adaptation: the role of educational and learning capital. In N. S. Phillipson, H. Stoeger & A. Ziegler (Eds.), Exceptionality in East Asia (pp. 18-39). London: Routledge.

Ziegler, A., Grassinger, R., Stoeger, H., & Harder, B. (2012). Das Beratungskonzept der Landesweiten Beratungs- und Forschungsstelle für Hochbegabung (LBFH) [The conseling concept of the State-Wide Research and Counseling Center for Giftedness]. In A. Ziegler, R. Grassinger & B. Harder (Eds.), *Konzepte der Hochbegabtenberatung in der Praxis [Conceptions of counseling the gifted put into practice]* (pp. 247-269). Münster: LIT.

Ziegler, A., & Heller, K. A. (2002). Conceptions of giftedness from a meta-theoretical perspective. In K. A. Heller, F. J. Mönks, R. J. Sternberg & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (2nd ed., pp. 3-21). Oxford: Elsevier Science Ltd.

Ziegler, A., Stoeger, H., Harder, B., & Balestrini, D. P. (2013). Gifted education in German-speaking Europe. *Journal for the Education of the Gifted, 36*(3), 384-411. doi: 10.1177/0162353213492247

Ziegler, A., Stoeger, H., Harder, B., Park, K., Portešová, Š., & Porath, M. (2014). Gender differences in mathematics and science: The role of the actiotope in determining individuals' achievements and confidence in their own abilities. *High Ability Studies, 25*(1), 35-51.

Table 1.

Rasch reliabilities and separation indices for all scales.

	Scale	Person			Item		
		N	*REL*[a]	*SEP*[b]	*N*	*REL*[a]	*SEP*[b]
Performance	German 1	347	0.82	2.11	36	0.97	5.98
	German 2	341	0.77	1.85~	36	0.97	6.12
	German 3	299	0.84	2.27	38	0.99	8.31
	English 1	331	0.95	4.23	178	0.99	9.55
	English 2	337	0.92	3.39	186	0.99	9.79
	English 3	284	0.94	4.03	187	0.99	8.75
	Mathematics 1	331	0.77	1.85~	32	0.99	8.70
	Mathematics 2	314	0.89	2.87	45	0.98	7.30
	Mathematics 3	285	0.89	2.81	44	0.98	6.82
	Verbal IQ 1	337	0.79	1.95~	45	0.97	5.93
	Numeric IQ 1	337	0.75	1.73~	45	0.99	8.26
	Nonverbal-IQ 1	337	0.80	1.99~	50	0.94	3.89
Motivation	Academic self-concept 1	345	0.77	1.84~	11	0.85	2.40
	Academic self-concept 2	319	0.81	2.07	11	0.85	2.38
	Mastery goal 1	345	0.61~	1.26~	8	0.93	3.56
	Mastery goal 2	319	0.73	1.64~	8	0.93	3.70
	Performance goal 1	345	0.91	3.21	15	0.98	7.38
	Performance goal 2	319	0.92	3.44	15	0.96	4.66
	Interest 1	345	0.74	1.68~	9	0.95	4.18
	Interest 2	319	0.80	1.97~	9	0.95	4.16
Learning behavior	Goal commitment 1	325	0.88	2.71	13	0.96	4.69
	Goal commitment 2	327	0.88	2.70	13	0.90	2.94
	Effort 1	325	0.78	1.91~	10	0.95	4.16
	Effort 2	327	0.81	2.10	10	0.91	3.21
	Concentration 1	325	0.89	2.86	14	0.97	5.82
	Concentration 2	327	0.89	2.86	14	0.96	4.68
	Working attitude 1	325	0.83	2.18	26	0.94	3.96
	Working attitude 2	327	0.86	2.47	26	0.95	4.21
School/class climate	Warmth 1	319	0.86	2.44	9	0.96	4.93
	Warmth 2	311	0.88	2.68	9	0.95	4.32
	Social and learning pressure 1	325	0.86	2.50	14	0.99	9.89
	Social and learning pressure 2	311	0.87	2.63	14	0.99	8.44
	Student centeredness 1	325	0.89	2.82	13	0.97	5.75
	Student centeredness 2	311	0.90	2.99	13	0.94	3.82
	Learning climate 1	324	0.77	1.84~	8	0.99	10.24
	Learning climate 2	311	0.80	1.99~	8	0.99	10.23
Actiotope	Subjective action space 1	310	0.75	1.74~	4	0.97	5.93
	Subjective action space 2	301	0.76	1.78~	4	0.96	4.65
	Goals 1	310	0.72	1.62~	5	0.98	6.90
	Goals 2	301	0.75	1.71~	5	0.98	8.09
	Environment 1	310	0.54~	1.09~	5	0.95	4.29
	Environment 2	301	0.47~	0.94~	5	0.98	6.61
	Anticipation 1	310	0.79	1.94~	5	0.99	8.85
	Anticipation 2	301	0.77	1.81~	5	0.98	7.67
	Applicability 1	310	0.77	1.81~	5	0.82	2.14
	Applicability 2	301	0.78	1.89~	5	0.91	3.13
	Feedback 1	310	0.84	2.25	4	0.97	5.34
	Feedback 2	301	0.85	2.37	4	0.97	5.94
	Action variations 1	310	0.83	2.18	5	0.95	4.27
	Action variations 2	301	0.86	2.43	5	0.95	4.59
	Correctness 1	310	0.78	1.88~	5	0.92	3.34
	Correctness 2	301	0.74	1.69~	5	0.83	2.21
	Modifiability 1	310	0.71	1.57~	5	0.98	6.35
	Modifiability 2	301	0.73	1.64~	5	0.96	4.74
	Stability 1	310	0.78	1.87~	4	0.91	3.22
	Stability 2	301	0.81	2.07	4	0.93	3.65

Note to Table 1.

Reliabilities and separation indices are calculated for persons and items where person reliability corresponds to classical reliability measures like Cronbach's alpha.

[a] Rasch reliability; indicates the ratio of true variance to observed variance which should exceed 0.70 but can be smaller when few items with few response categories are used (Bond & Fox, 2007);

[b] Rasch separation index; reports the ratio of true variance to error variance in the Rasch measures (signal to noise ratio) which should exceed the value of 2 i.e., a distance of at least two root mean standard errors between neighboring items or persons allowing their differentiation;

~ measures below the critical value;

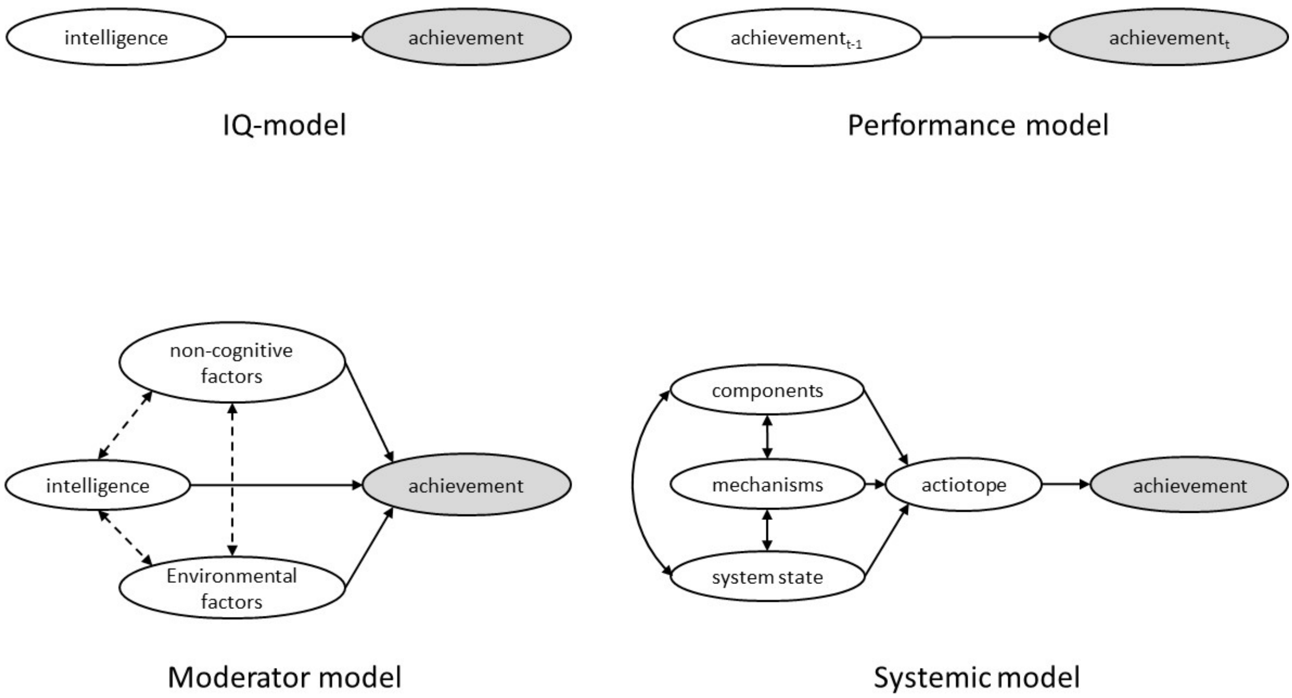

Figure 1. Overview of the four models of giftedness and expertise put to the empirical test.

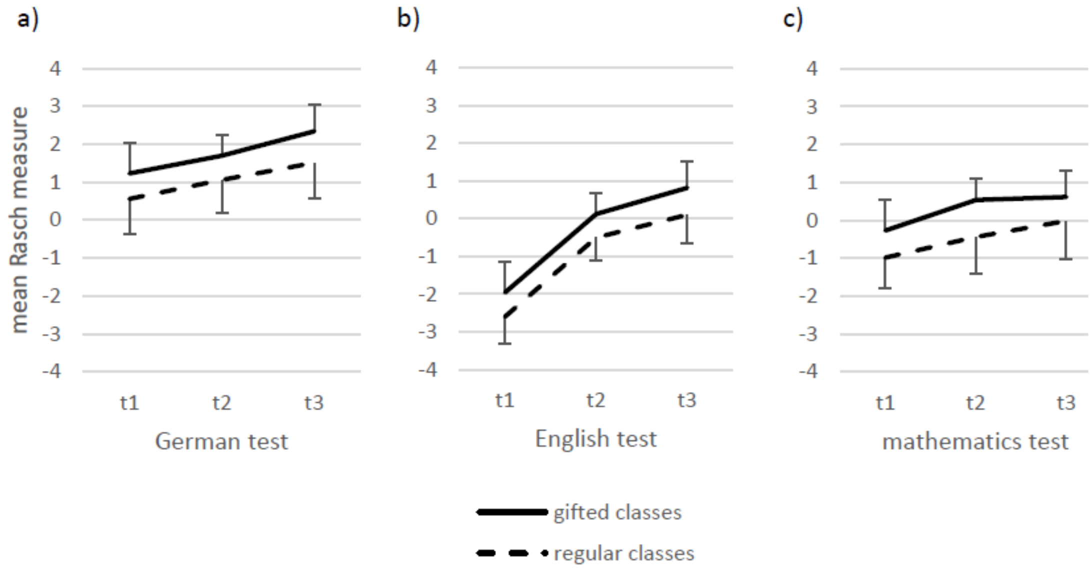

Figure 2. Means (± 1 *SD*) of a) German, b) English and c) mathematics achievement of the gifted and regular classes from t1 to t3. Achievement was measured by standardized, Rasch-analyzed and linked tests.

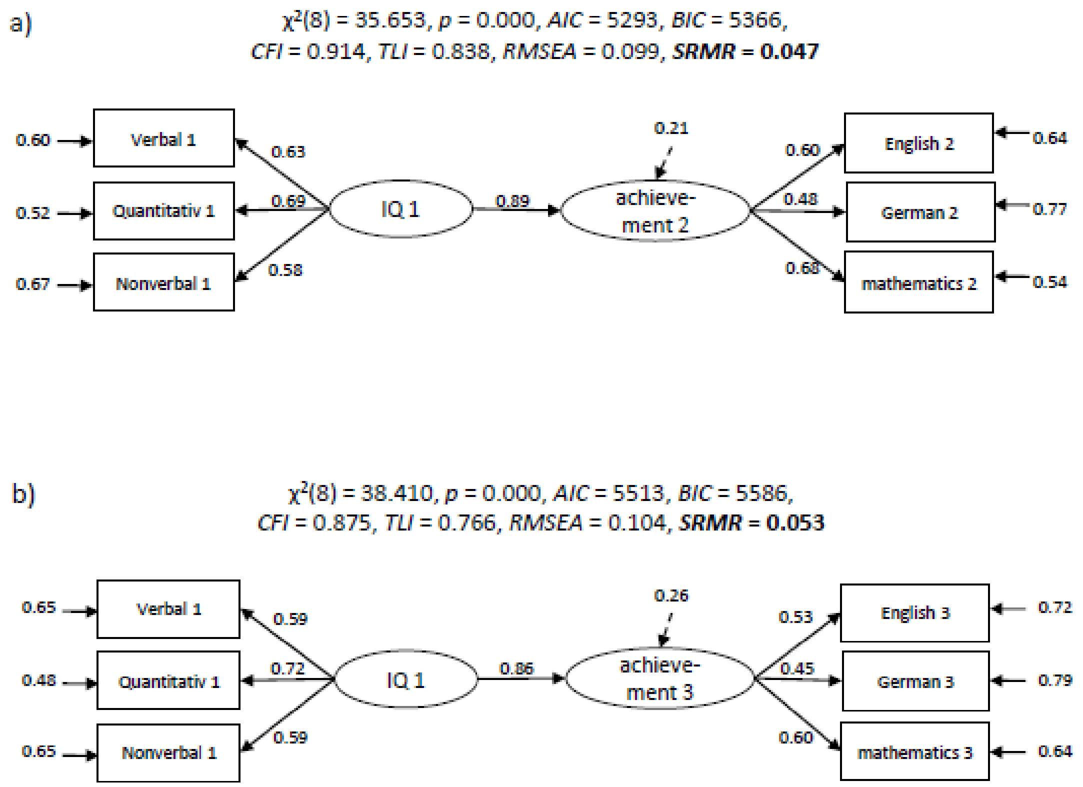

Figure 3. Original IQ-model. Predictions of school achievement at t2 (a) and t3 (b) by general intelligence measured at t1. Fit indices in bold face meet the criteria of acceptable fit outlined before, solid arrows represent significant, broken lined arrows non-significant standardized coefficients, factor loadings, or residuals.

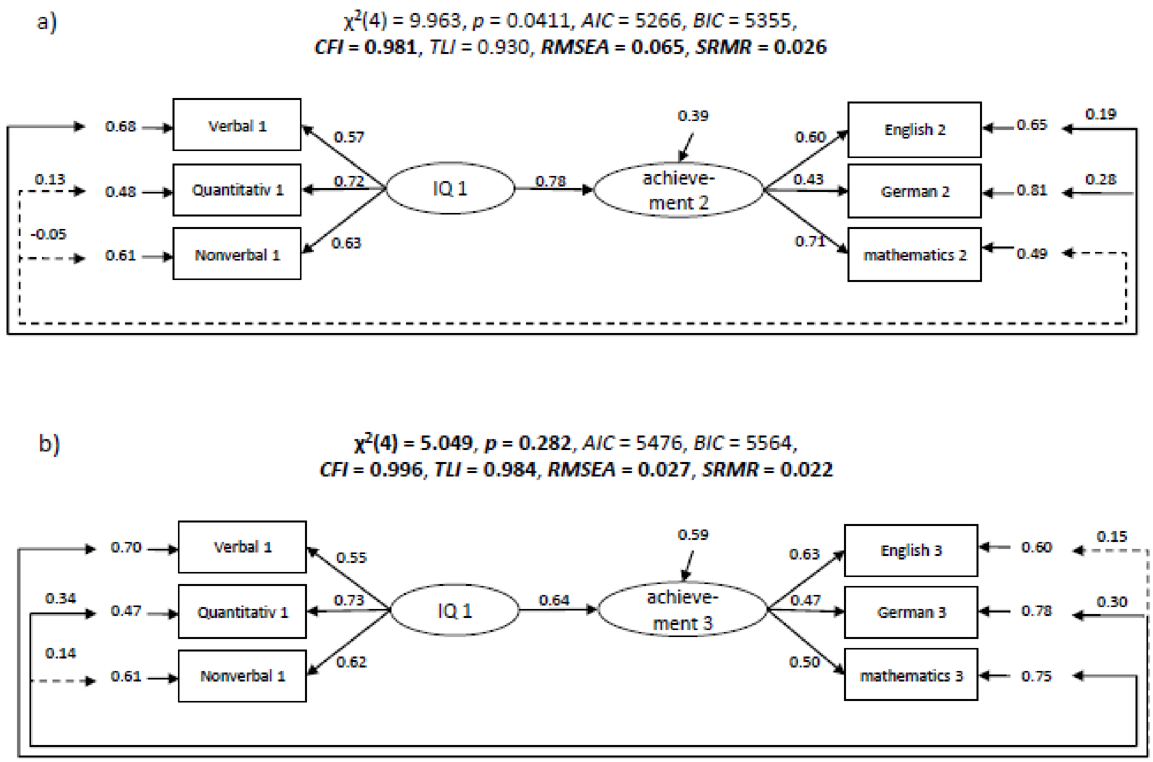

Figure 4. Modified IQ-model. Prediction of achievement at t2 (a) and t3 (b) by general intelligence measured at t1 with additional domain specific correlations between residuals (legend see Figure 3).

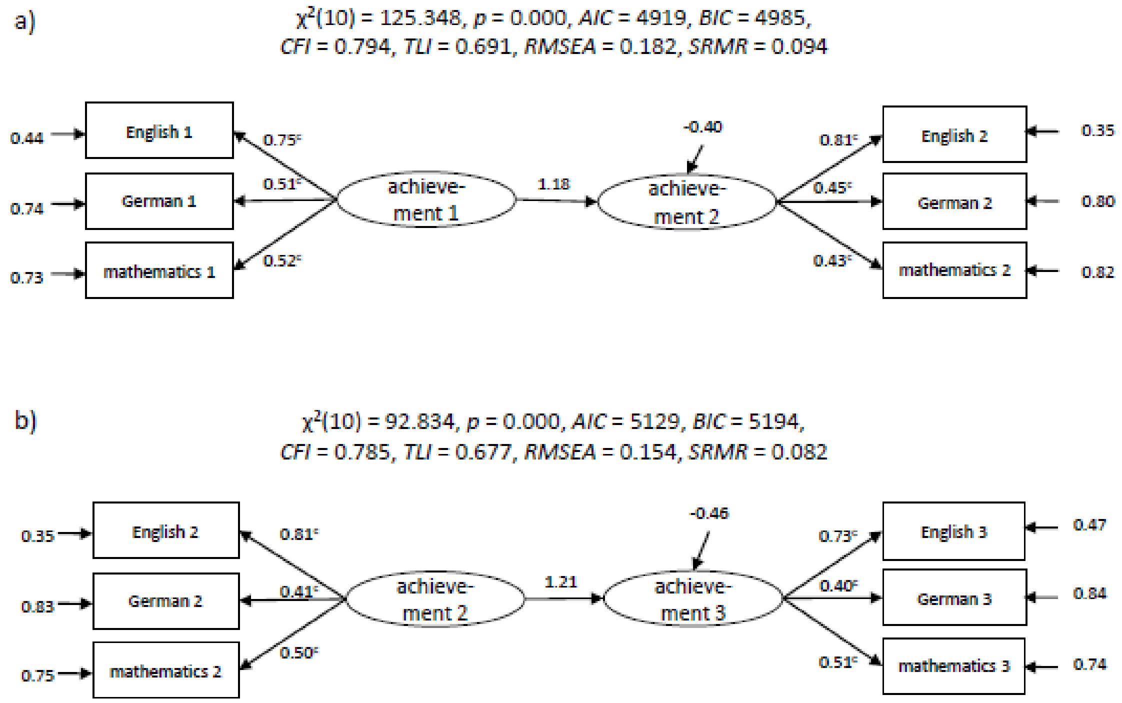

Figure 5. Original performance model. Predictions of school achievement at t2 (a) and t3 (b) by school achievement at the preceding measuring point (legend see Figure 3; “c” indicates that unstandardized factor loadings were held constant between the measuring points to gain equivalent indicators in the CFAs and to ensure comparability of the latent factors. Displayed standardized factor loadings differ again).

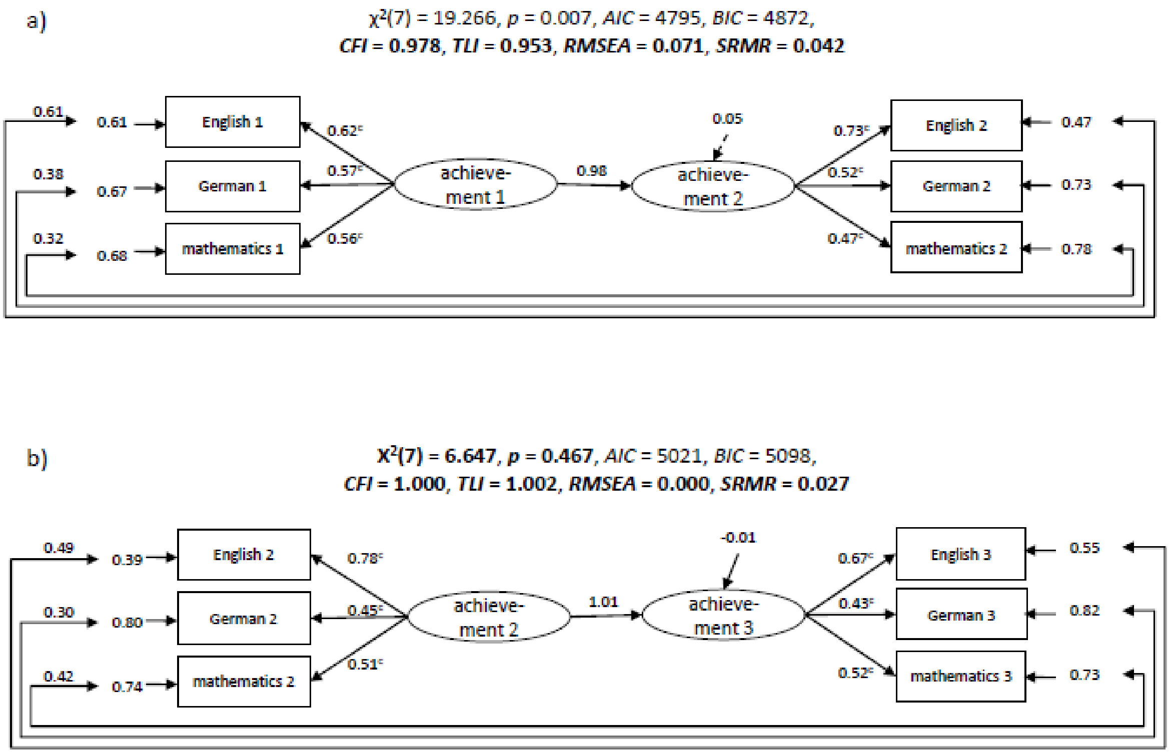

Figure 6. Modified performance model. Prediction of achievement at t2 (a) and t3 (b) by school achievement at the preceding measuring point with additional within-domain correlations between residuals (legend see Figure 3; "c" indicates that unstandardized factor loadings were held constant between measuring points to gain equivalent indicators in the CFAs and to ensure comparability of the latent factors. Displayed standardized factor loadings differ again).

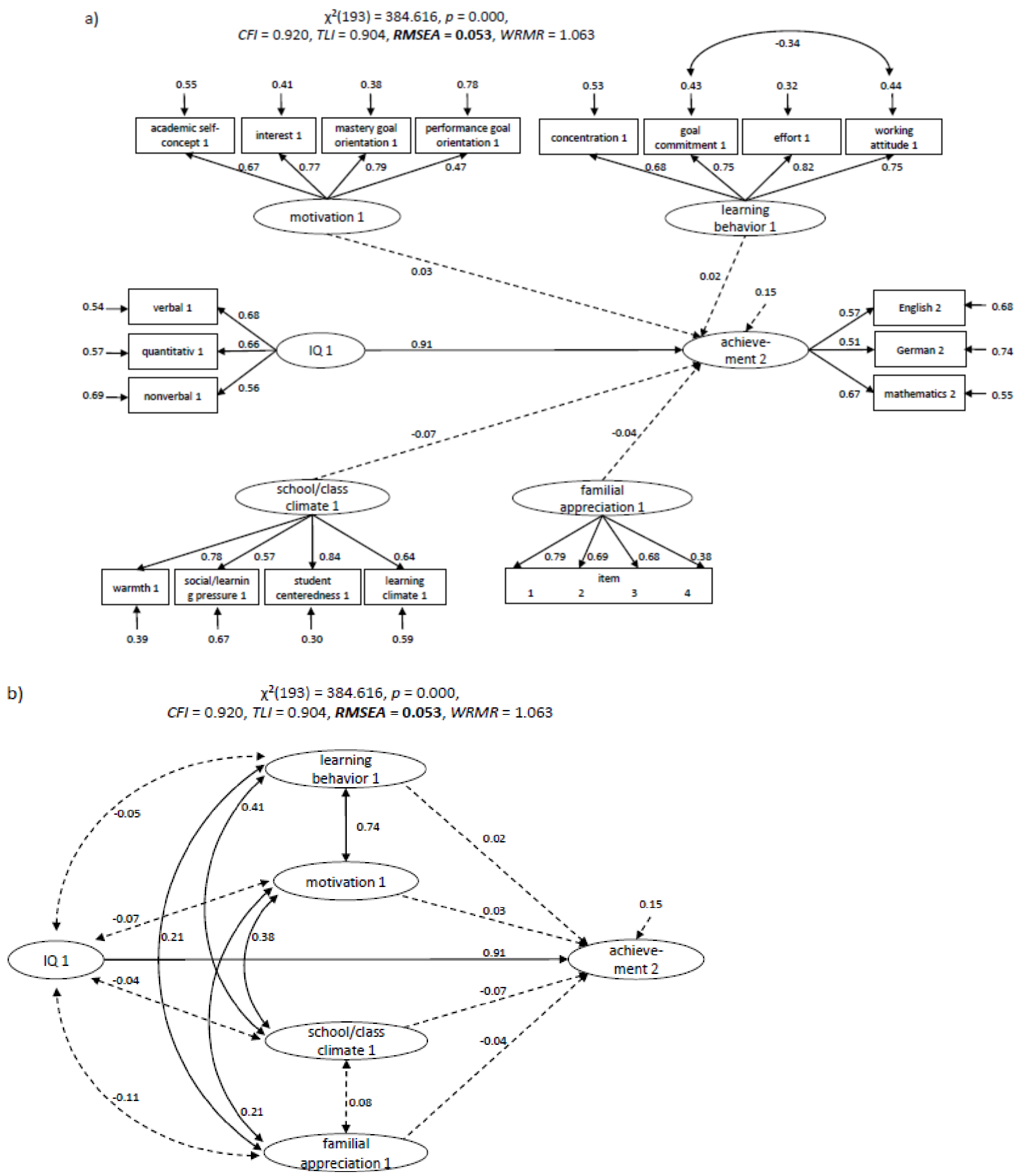

Figure 7. Moderator model for the prediction of achievement at t2. Achievement is regressed on intelligence and two non-cognitive personal moderators (learning behavior and motivation) as well as two environmental moderators (school/class climate and appreciation of school matters in the family) displayed in figure 7a. Figure 7b separately pictures the inter-correlations between the moderators of the model (Legend see Figure 3).

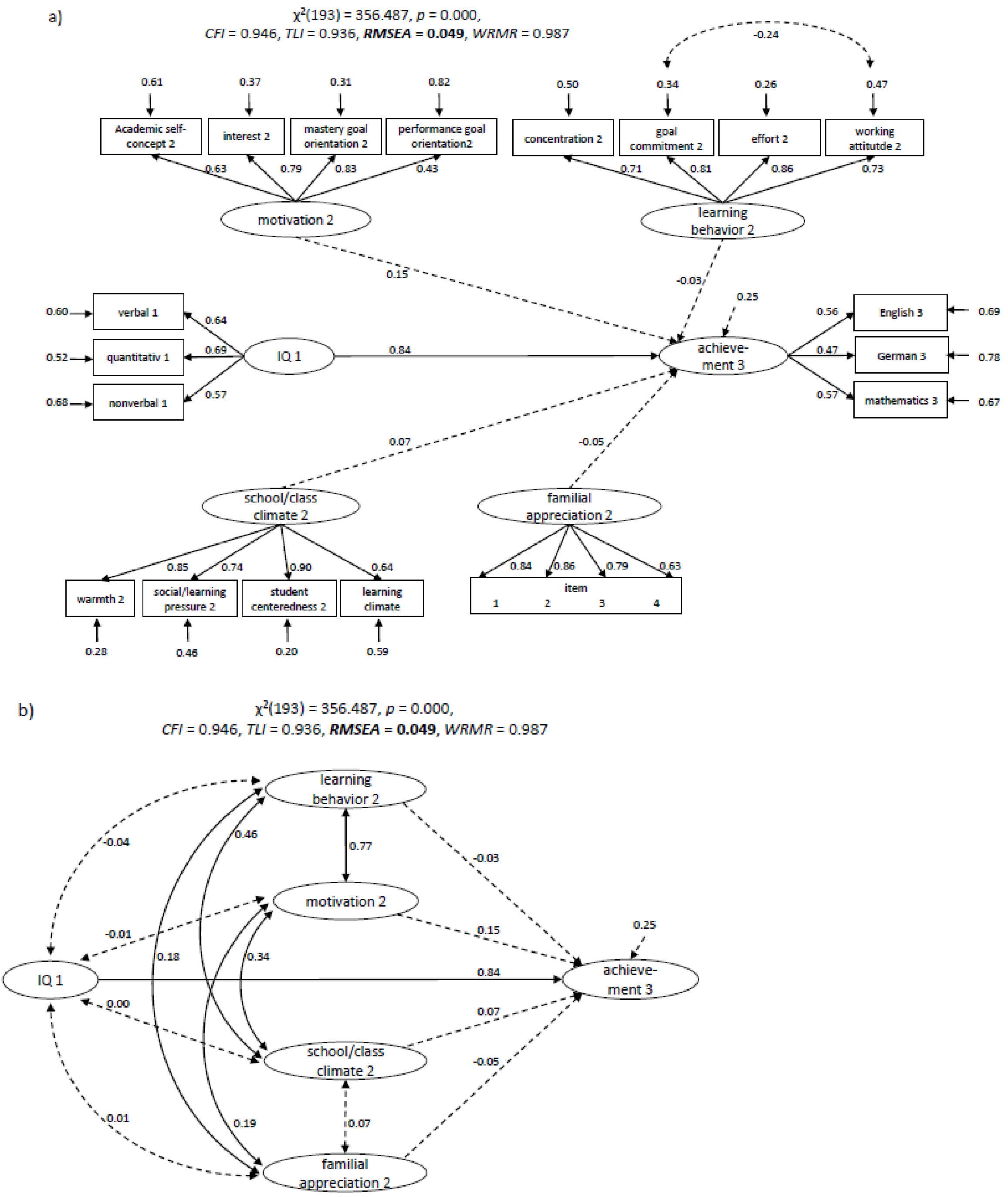

Figure 8. Moderator model for the prediction of achievement at t3. Achievement is regressed on intelligence and two non-cognitive personal moderators (learning behavior and motivation) as well as two environmental moderators (school/class climate and appreciation of school matters in the family) displayed in figure 8a. Figure 8b separately pictures the inter-correlations between the moderators of the model (Legend see Figure 3).

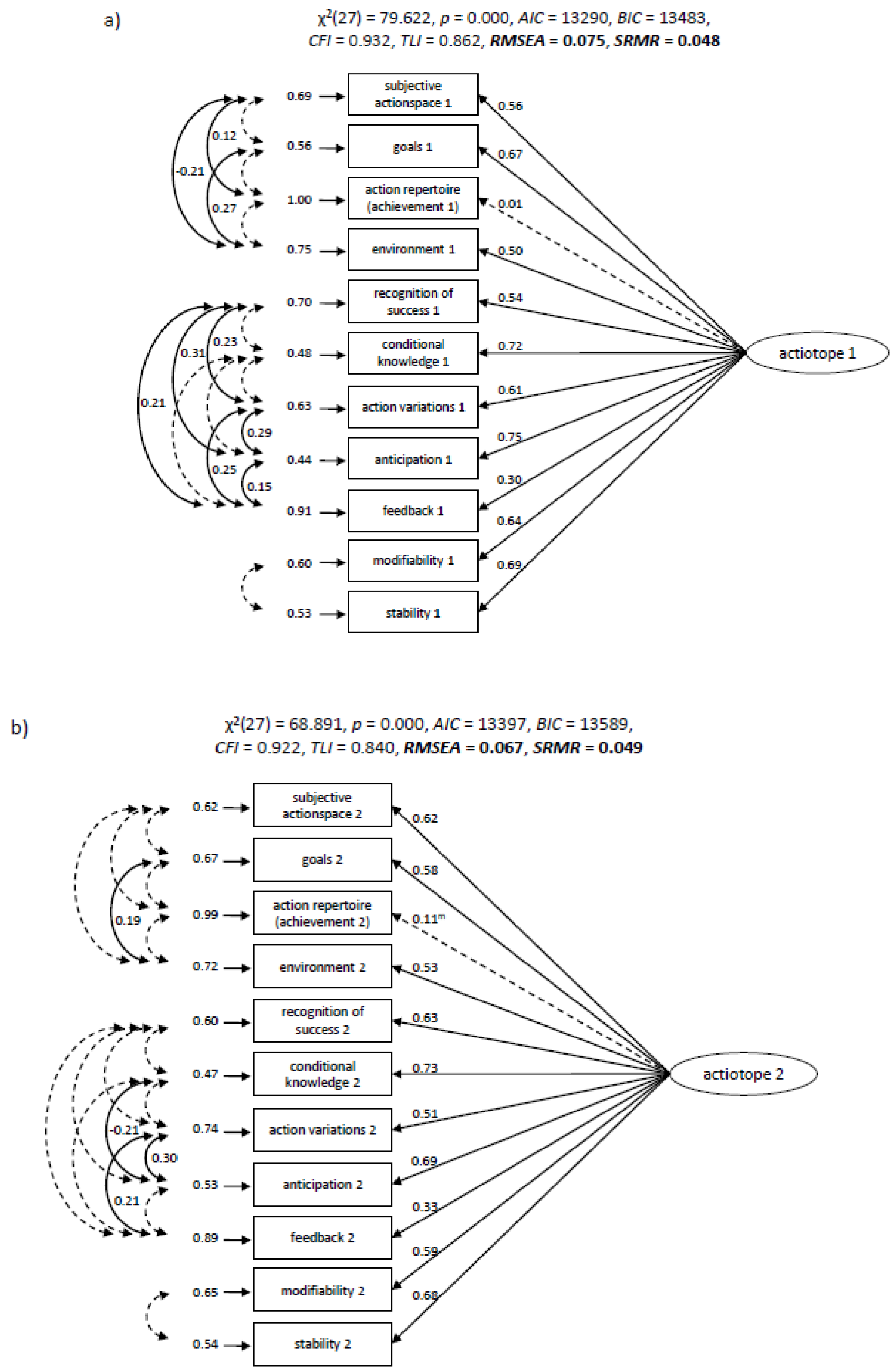

Figure 9. Model of the Actiotope's construct validity (systemic model). For t1 (a) and t2 (b) the Actiotope is modelled by the constituting constructs assigned to the categories of components (subjective action space to environment), mechanisms (recognition of success to feedback) and the system state (modifiability and stability). Interactions were modelled as residual correlations between the constructs of the same category. (Legend see Figure 3; "m" indicates marginal significance i.e., .05 < $p \leq$ .10).