Arens, A. Katrin; Jansen, Malte

Self-concepts in reading, writing, listening and speaking. A multidimensional and hierarchical structure and its generalizability across native and foreign languages

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
The journal of educational psychology 108 (2016) 5, S. 646-664, 10.1037/edu0000081

Self-concepts in reading, writing, listening, and speaking: A multidimensional and hierarchical structure and its generalizability across native and foreign languages

A. Katrin Arens
Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), German Institute for International Educational Research, Germany
&
Malte Jansen
Institute for Educational Quality Improvement, Humboldt-Universität zu Berlin, Germany and German Institute for International Educational Research, Germany

Author Note
A. Katrin Arens, Centre for Research on Education and Human Development and Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), German Institute for International Educational Research, Frankfurt, Germany; Malte Jansen, Institute for Educational Quality Improvement, Humboldt-Universität zu Berlin, Berlin, Germany and Department of Educational Governance, German Institute for International Educational Research, Berlin, Germany. Correspondence concerning this article should be addressed to A. Katrin Arens, Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), German Institute for International Educational Research, Schloßstraße 29, D-60486 Frankfurt am Main, Germany, Telephone: + 49 69 24708 138, Email: arens@dipf.de

Acknowledgements
We would like to thank the Institute for Educational Quality Improvement (IQB) in Berlin for the gratuitous allowance of data access. More specifically, we want to thank Ulrich Schroeders for comments on a previous version of this manuscript and Stefan Schipolowski for helpful discussions with regard to the operationalization.

This is the prepublication version of the following manuscript:

Arens, A.K., & Jansen, M. (2016). Self-concepts in reading, writing, listening, and speaking: A multidimensional and hierarchical structure and its generalizability across native and foreign languages. *Journal of Educational Psychology, 108*, 646–664. http://doi.org/10.1037/edu0000081

This paper is not the copy of record and may not exactly replicate the authoritative document published in Journal of Educational Psychology.

Abstract

Academic self-concept has been conceptualized as a multidimensional and hierarchical construct. Previous research has mostly focused on its multidimensionality distinguishing between verbal and mathematical self-concept domains and only few studies have examined the factorial structure within specific self-concept domains. The present study aims to extend the scarce evidence of the simultaneous operation of multidimensionality and hierarchy within language self-concepts. Skill-specific (i.e., reading, listening, speaking, and writing) and global self-concept measures for German as students' native language and English and French as students' foreign languages were administered to a sample of 4,257 ninth-grade German students. Confirmatory factor analyses demonstrated better levels of fit for multidimensional models with highly correlated but separate factors for each skill-specific self-concept compared to unidimensional models. The skill-specific facets could be aggregated into a higher-order factor for each language, demonstrating a hierarchical structure. Differential gender effects on the various skill-specific self-concept facets provided further evidence of the multidimensional nature of language self-concepts although differential relations to skill-specific achievement measures were not found. The high correlation between the higher-order and global self-concept factors and similar gender effects and relations to achievement supported the notion of the similarity of both constructs. A comparable pattern of results emerged across the three languages of German, English, and French, but detailed inspection indicated that multidimensionality seems to be more pronounced in the native language domain (German). The findings of the study are discussed regarding their implications for the assessment of language self-concepts and future research on the structure of language self-concepts.

Keywords: academic self-concept; multidimensionality; hierarchy; languages

Self-concepts in reading, writing, listening, and speaking: A multidimensional and hierarchical structure and its generalizability across native and foreign languages

Classical self-concept theories such as the seminal model proposed by Shavelson, Hubner, and Stanton (1976) conceptualized self-concept as a multidimensional and hierarchical construct. It has thus been assumed that self-concept can be distinguished into specific self-concept facets relating to different domains (multidimensionality) and that these specific self-concept facets can be aggregated to broader self-concept factors (hierarchy). Research on the structure of academic self-concept has however primarily focused only on the multidimensionality of self-concept. This is also reflected in research on verbal self-concept which has indeed been found to be distinct from math self-concept (Marsh, 1986, 1990; Möller, Pohlmann, Köller, & Marsh, 2009), and to consist of separable self-concept facets for different languages (Marsh, Kong, & Hau, 2001; Marsh & Yeung, 2001; Möller, Streblow, Pohlmann, & Köller, 2006; Xu et al., 2013; Yeung & Wong, 2004). However, only two studies (Lau, Yeung, Jin, & Low, 1999; Yeung et al., 2000) examined the structure within specific language self-concepts. The results of these studies suggest the simultaneous operation of multidimensionality and hierarchy within self-concepts related to specific languages. In essence, students were found to hold separate self-concepts for specific verbal skills (e.g., reading, writing) suggesting multidimensionality, which, however, were highly related and could thus be integrated into a higher-order factor also indicating hierarchy. The present study aims to expand the so far scarce evidence of multidimensionality and hierarchy that are at play within language self-concepts by further extending the applied approaches. Beyond examining the within-structure of specific language self-concepts, we consider self-concept–achievement relations and gender differences as well as the generalizability of findings across students' native and foreign languages. For this purpose, we measured the self-concepts of ninth-grade German secondary school students related to their native language (German) and two foreign languages (English, French) with regard to four verbal skills (reading, writing, listening, and speaking).

The Construct of Academic Self-Concept

Academic self-concept has become a popular construct in educational psychology due to its substantial relations to desirable educational outcomes including achievement (Chen, Yeh, Hwang, & Lin, 2013; Hattie, 2008; Huang, 2011; Marsh & Martin, 2011; Marsh & O'Mara, 2008; Valentine, DuBois, & Cooper, 2004), motivation (Nagengast et al., 2011; Wigfield & Eccles, 2000), effort (Trautwein, Lüdtke, Schnyder, & Niggli, 2006), and educational choices (Marsh & Yeung, 1997; Parker, Marsh, Ciarrochi, Marshall, &

Abduljabbar, 2014). The structural model proposed by Shavelson and colleagues (1976) marked the beginning of sophisticated and empirically based self-concept research. In this model, multidimensionality and hierarchy were defined as the core characteristics of self-concept. Accordingly, students form self-concepts in different school subjects (multidimensionality) which can then be aggregated to a global academic self-concept (hierarchy). However, math and verbal self-concepts have been consistently found to be almost uncorrelated (Marsh, 1986, 1990; Möller et al., 2009) which is inconsistent with the assumption of a global academic self-concept. Thus, from a between-domain perspective (i.e., across math and verbal domains), multidimensionality prevails in the structure of self-concept while there is little evidence of hierarchy although it has become the focus of more recent considerations (e.g., Brunner et al., 2010).

The Structure of Language Self-concepts

Within the verbal domain, separate self-concepts have been found for different languages such as students' native and foreign languages (Marsh et al., 2001; Marsh & Yeung, 2001; Möller et al., 2006; Xu et al., 2013), for different Chinese dialects (Yeung & Wong (2004) or for two different foreign languages (Brunner et al., 2010). Thus, when focusing on self-concepts for specific languages (i.e., on the structure *between* language self-concepts), there is little evidence of hierarchy (i.e., no support for a global verbal self-concept) but strong evidence of multidimensionality (i.e., separate self-concepts for different languages). When only considering the structure *within* self-concepts for specific languages, research indicates a new debate between the relative strengths of multidimensionality and hierarchy.

Multidimensionality within Language self-concepts

Multidimensionality within specific language self-concepts would mean that it is possible to distinguish various specific self-concepts related to verbal skills. In fact, language performance includes at least the four skills of reading, writing, speaking, and listening which have been conceptualized as different language systems (Berninger, 2000) referring to different parts of the body (ear, mouth, eye, and hand), producing different outputs, showing differential developmental trajectories, and depending on different inputs. In research and practice of language assessment (Buck, 2001; Harris, 1969; Lado, 1961; Rupp, Vock, Harsch, & Köller, 2008), the distinction among these four verbal skills of reading, writing, speaking, and listening has been well established as these facets include all combinations of productive and receptive as well as oral and written skills (reading: receptive written, writing: productive written, speaking: productive oral, listening: receptive oral). For example, the two most

popular language assessment tests in English, the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS) include these four skills (e.g., Sawaki, Stricker, & Oranje, 2008). Furthermore, both the framework for foreign language testing in the National Assessment of Educational Progress (NAEP) established in the US (National Center for Educational Statistics, 2015) and the German national educational standards for German and the foreign languages English and French refer to these four skills (Köller, Knigge, & Tesch, 2010; Rupp et al., 2008). In Germany, the national educational standards provide the basis for all school curricula indicating that the distinction among these verbal skills plays an important role for the planning and orchestration of classroom instruction. Therefore, students might display separate self-concepts as the subjective representations of these skills.

Indeed, there has been empirical evidence for the construct of reading self-concept (Aunola, Leskinen, Onatsu-Arvilommi, & Nurmi, 2002; Chapman & Tunmer, 1995; see also Wigfield & Guthrie, 1997) which has been found to be separable from math self-concept (Arens, Yeung, & Hasselhorn, 2014), to display substantial relations to reading-related outcomes (e.g., reading achievement or home literacy environment, Chapman & Tunmer, 1995; Katzir, Lesaux, & Kim, 2009), and to be subject to stereotypical gender differences in favor of girls (Eccles, Wigfield, Harold, & Blumenfeld, 1993). Less research, in contrast, has been conducted regarding the presumptive existence of self-concepts related to the other verbal skills (i.e., listening, speaking, and writing, but see Pajares and Valiante (2001) for a study on writing self-concept). Yet, in the study of Lau et al. (1999) examining the factor structure of self-concept in English as a foreign language with Hong Kong university students, self-concepts related to writing, reading, listening, and speaking emerged as separate well-defined factors (see also Yeung et al.. 2000).

Hierarchy within Language Self-concepts

Even in the case of the existence of separate self-concepts related to different verbal skills (multidimensionality), their interrelations remain questionable. Students' performances on the different verbal skills of reading, writing, speaking, and listening were found to be substantially associated to each other (Abbott, Berninger, & Fayol, 2010; Ehri, 2000). In addition, even though the different skills are subject to unique mental processes, they seem to share a common underlying core of mental processing (Berninger, 2000). Therefore, self-concept facets corresponding to the four skills should be highly interrelated. In this case, instead of arguing for the existence of separate skill-specific language self-concepts implying multidimensionality, it may be rather assumed that multiple self-concepts related to different

verbal skills can be aggregated to a global language self-concept factor implying hierarchy. This assumption is further substantiated by the fact that students commonly only obtain one single school grade representing their performances in a language domain but rarely get differential feedback on the different verbal skills which would facilitate the establishment of separate skill-specific self-concepts.

The relations found between different verbal skills, however, are not perfect so that their subjective representations within students may still enable the formation of separate skill-specific self-concepts. For example, in the study of Retelsdorf and Köller (2014), the cross-sectional correlations between reading comprehension and spelling in German as students' native language were only $r = .52$ and $r = .56$ in samples of fifth- and seventh-grade students. In addition, differential implicit feedback (e.g., a teacher's comment that a good school grade on a writing exam might compensate for inferior oral performance) and different instructional methods for teaching these skills might also result in separable self-concepts for various verbal skills which cannot be adequately represented by a global factor.

The Co-Existence of Multidimensionality and Hierarchy

In order to remedy this controversial juxtaposition of multidimensionality and hierarchy within language self-concepts, the findings of two previous studies (Lau et al., 1999; Yeung et al., 2000) have argued that separability of multiple skill-specific facets (multidimensionality) and the assumption of a global factor (hierarchy) are not mutually exclusive. Rather, the results of these studies could provide evidence for the *co-existence* of multidimensionality and hierarchy. Lau et al. (1999) showed that self-concepts related to writing, reading, listening, and speaking in English formed separate well-defined factors (implying multidimensionality) which were, however, found to be substantially positively interrelated and could be integrated into a higher-order English self-concept factor (implicating hierarchy). Yeung et al. (2000) demonstrated a similar pattern of results for ninth-grade Australian students' self-concepts in English and languages other than English (LOTE). In essence, in both the domains of English and LOTE, students were found to display separate but substantially related factors for reading, writing, and speaking self-concepts which could be integrated into a higher-order factor.

In both studies, the notion of hierarchy was further substantiated by investigating the correlation between a higher-order factor and a factor for global self-concept. Commonly applied in self-concept measurement (e.g., Arens, Yeung, Craven, & Hasselhorn, 2013; Wigfield et al., 1997), global self-concept scales consist of items directly asking for students' self-evaluations in a language domain without referring to any specific skills within this

domain (e.g., "I am good at English"). Thus, there seems to be two alternative ways of gaining insight into students' overall self-perception related to a specific language domain. First, skill-specific self-concepts can be assessed which then are used to form a higher-order factor. Second, as a more parsimonious way, the global self-concept related to a language domain can be assessed directly. In the study of Lau et al. (1999) the higher-order factor built on the basis of self-concepts for writing, reading, listening, and speaking in English was found to be perfectly related (r = .97) to an independently assessed global English self-concept factor. This finding argues for a strong hierarchical structure within English self-concept in which the higher-order and global factors constitute equivalent constructs. In the study of Yeung et al. (1999), the higher-order factor also showed a significant positive relation to an independently measured global self-concept factor both within the domains of English and LOTE. However, evidence of the equivalence of the higher-order and global factors was much weaker here as their correlations were only r = .48 for English and r = .65 for LOTE. Thus, this study more strongly argues for the discriminability instead of equivalence of higher-order and global language self-concepts, therefore weakening the argument for a predominantly hierarchical structure within language self-concepts.

Further Testing of the Within-Structure of Language Self-concepts

Even though the co-existence of multidimensionality and hierarchy within language self-concepts was demonstrated by Lau et al. (1999) and Yeung et al. (2000), the relative extent of these two features (i.e., multidimensionality and hierarchy) is yet to be further examined. The analytical approach in the studies of Lau et al. (1999) and Yeung et al. (2000) was restricted to the investigation of the factor structure of language self-concepts. The present study aims to go beyond this approach by additionally examining achievement relations and gender differences.

Achievement relations. Academic self-concept has generally been found to be related to academic achievement whereby this relation follows a domain-specific pattern. Thus, the highest relations between self-concept and achievement occur when both constructs address the same content domain and are located on the same level of hierarchy (Huang, 2011; Marsh & Craven, 2006; Swann, Chang-Schneider, & Larsen McClarty, 2007; Valentine et al., 2004). In consequence, to substantiate multidimensionality within language self-concepts, skill-specific self-concept facets are expected to show higher relations to skill-specific achievement measures (e.g., reading self-concept and reading achievement) than to other achievement measures (e.g., reading self-concept and listening achievement).

The inspection of achievement relations is also helpful with respect to insights into the similarity versus differentiability of higher-order and global language self-concept factors and thus into the interpretation of the higher-order factor. If the higher-order and global factors demonstrated similar relations to achievement, this would argue for their equivalence and thus for hierarchy within language self-concepts. Differential achievement relations of the higher-order and global factors in turn would imply that both depict separate constructs associated with differential predictive value for achievement. In addition, this finding would indicate multidimensionality within language self-concepts, i.e., a higher-order factor composed of skill-specific language self-concepts entails a meaning that is different from that of a global language self-concept.

Gender differences. Numerous studies have investigated gender differences in self-concept facets finding that gender differences generally follow gender stereotypes with girls displaying higher mean levels of verbal self-concept than boys (Wilgenbusch & Merrell, 1999), considering both the global verbal domain (Arens et al., 2013; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Marsh, 1989; Marsh & Yeung, 1998a; Skaalvik & Rankin, 1990) and the skill-specific domain of reading (Marsh & Ayotte, 2003; Wigfield et al., 1997). The examination of gender differences might contribute to validating the multidimensional structure that potentially operates within specific language self-concepts. Differential gender effects on the various skill-specific self-concept facets (e.g., higher mean levels for girls on reading self-concept but higher mean levels for boys on listening self-concept) would clearly support the differentiability of skill-specific self-concepts within language self-concepts. Furthermore, the investigation of gender effects might help clarify the meaning of a higher-order factor relative to that of a global factor for language self-concepts. Comparable gender effects for the higher-order and global factors would argue for the similarity of both constructs implicating strong hierarchy within the structure of language self-concepts whereas differential gender effects would indicate the dissimilarity of both constructs arguing for multidimensionality within language self-concepts.

Generalizability across Languages

Although previous findings have provided evidence for the co-existence of multidimensionality and hierarchy within self-concepts for specific languages (Lau et al., 1999; Yeung et al., 2000), it has remained unclear whether this structural feature might or might not differ across languages including native and foreign languages. The findings of Yeung et al. (2000) show that both the self-concepts for English as a native language and for LOTE display a multidimensional and hierarchical structure since skill-specific self-concept

facets were found which could be integrated into a higher-order factor. However, subtle differences became apparent when considering the correlations between higher-order and global self-concept factors which was higher (r = .65) for LOTE than for English (r = .48). These findings imply that students discriminate more strictly among skills in their native language than among foreign language skills, leading to a weaker hierarchy but stronger multidimensionality within native language self-concept. Foreign languages are learned as school subjects and specific skills such as reading, writing, listening, and speaking are taught at the same time. Students are assigned a common grade so that they might perceive specific skills in foreign languages to be closely related to each other. With regard to native language, by contrast, students obtain a single grade for their achievement in the school subject related to native language, but other school subjects and various life domains also require separate specific native language skills (e.g., reading instructions and textbooks, writing letters, listening to the news, etc.). Therefore, students might rather receive differential feedback regarding their native language skills which might lead to more differentiated (i.e., less correlated) self-concepts for native language skills. In addition, students acquire native language skills consecutively (writing and reading mostly at school but listening and speaking earlier on) also possibly resulting in differential learning experiences and thus separate skill-specific native language self-concepts.

Although these arguments could support the conjecture that multidimensionality (i.e., the differentiation into skill-specific facets) might be more pronounced for self-concepts related to students' native language than for self-concepts related to foreign languages, more research on this issue is needed. In the study of Yeung et al. (2000), LOTE covered a wide range of language domains (e.g., Arab, Vietnamese, Chinese) without further distinguishing between languages. Additionally, LOTE included both languages spoken at home and languages studied at school without further differentiating between these two categories.

The Present Study

Previous research has provided first evidence of a coexisting multidimensional and hierarchical structure within specific language self-concepts (Lau et al., 1999; Yeung et al., 2000). The present study aims to further substantiate this assumption by providing more and new empirical evidence. To substantiate multidimensionality within language self-concepts, we first test whether students differentiate between skill-specific self-concepts related to reading, listening, writing, and speaking. To account for the presumptive hierarchical structure, we test the adequacy of a higher-order factor model in which a higher-order factor is derived from the skills-specific self-concepts. We then use two approaches going beyond

previous research (Lau et al., 1999; Yeung et al., 2000) to further examine the expected multidimensionality and hierarchy within language self-concepts. First, we study relations to achievement measures. In this context, both school grades and standardized achievement test scores are utilized as achievement indicators. The allocation of school grades is based on a classroom-specific frame of reference ("grading on a curve") making it difficult to compare school grades across classes, schools, and school types. By contrast, standardized achievement test scores are comparable across classrooms but have been found to be less highly related to students' self-concept than school grades. The strong relations between school grades and self-concept are due to the high salience, comparability, and regular disclosure of school grades as well as to their important implications for educational decisions (Marsh et al., 2014; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005). Therefore, school grades have been conceptualized to be the main source for students' self-concept formation. Given the different strengths and weaknesses of the two achievement indicators, studies combining school grades and standardized achievement test scores are preferable.

The present study further examines gender differences in language self-concepts. In this context, we consider the structure of language self-concepts to probe whether the presumed interplay between multidimensionality and hierarchy applies to both boys and girls. In addition, the inspection of gender differences in the mean levels of the various self-concept facets serves to explore the multidimensionality and hierarchy of language self-concepts. Multidimensionality within language self-concepts would be supported in the case of differential gender effects on the various skill-specific self-concept facets and in the case of differential gender effects on the higher-order and global factors. Similar gender effects on the higher-order and global factors would in turn argue for hierarchy within language self-concepts.

This study incorporates self-concept and achievement measures for German as students' native language, and English and French as students' foreign languages. Therefore, we further aimed to find out whether the interplay of multidimensionality and hierarchy within language self-concepts generalizes or varies across students' native language (German) and two foreign languages (English and French).

Method

Sample

The data analyzed in the present study are taken from a pilot study for the German National Assessment Study 2015. National assessment studies are conducted at regular

intervals to determine the extent to which a priori specified national educational standards are met in the 16 German federal states. The data were collected in spring 2013.

The sample of the present study ($N = 4{,}257$[1]) consists of ninth-grade German secondary school students. In line with expectations for this grade level, students' mean age was 14.89 years ($SD = 0.64$; range 13 to 18 years). The total sample can be divided into two subsamples. The first subsample (Sample 1; $N = 2{,}528$) including 1,250 (49.4%) boys and 1,278 (50.6%) girls learned English as their first foreign language. They completed achievement tests and self-concept measures related to both German and English. The second subsample (Sample 2; $N = 1{,}729$) with 652 (37.7%) boys, 1,076 (62.2%) girls, and one (0.1%) student without indicated gender learned French as their first and English as their second foreign language. The students of Sample 2 completed achievement measures in French and self-concept measures in German, English, and French. In Germany, the secondary school system comprises different academic tracks to which the students are allocated based on their performance in elementary school, parents' preferences, and elementary school teachers' recommendations. The first subsample comprises students of all academic tracks including the lower ability track [*Hauptschule*, $N = 212$ (8.4%)], the middle track [*Realschule*, $N = 552$ (21.8%)], mixed track school forms [i.e., *Gesamtschule, Sekundarschule, Mittelschule*, $N = 675$ (26.7 %)], and the academic track [*Gymasium*, $N = 1089$ (43.1%)]. The second subsample (French as a first foreign language) most of all includes students from the academic track [$N = 1557$ (90.1%)] but also a few students from the middle track [$N = 79$ (4.6%)] and mixed track schools [$N = 93$ (5.4 %)]. Here, the academic track is overrepresented because all students from this track are required to learn two foreign languages compulsorily whereas in the other secondary school tracks, most students only learn one foreign language (English).

Procedure

Within randomly selected schools (in previously defined federal states of Germany), the participating ninth-grade students were also randomly drawn without consideration of any background variables. For the students, participation in the achievement tests was mandatory. Participation in the student survey with the self-concept measures was voluntary and required parental consent. Yet, the participation rates for the student questionnaire were high (Sample 1: $N = 83\%$; Sample 2: $N = 87\%$). Hence, our sample is sufficiently representative in terms of students' socioeconomic background and migration status.

The data collection started with the achievement tests which were completed by the students in two consecutive sessions each lasting 60 min, with a 15 min break in between. The

[1] This sample size refers to the subsample of students who took part in both the (compulsory) achievement tests and the (voluntary) student questionnaire. The total sample of students drawn for participation was 5,033.

questionnaire including the self-concept measures was completed afterwards. Both the achievement tests and the self-concept questionnaire were administered by trained university students who utilized a manual containing the instructions during the data collection process to ensure comparable test conditions and objectivity. Although not involved in the data collection process itself and without access to the students' answers, a school official was present in most cases to safeguard students' discipline and cooperation. All students were informed about the anonymous and confidential treatment of their data.

Measures

Achievement. In the present study, we used both school grades and achievement test scores as achievement measures. School grades in German, English and French (if applicable) that students had obtained in the latest school reports were reported by the school officials. School grades reflect students' achievement in the corresponding language domains in a broad and skill-spanning way based upon their accomplishments in various exams or tests throughout the school year. In the German educational system, school grades range between 1 and 6 with 1 indicating the best grade. For ease of interpretation, school grades were reversely coded such that higher values indicate higher achievement.

In contrast to school grades as a skill-spanning achievement measure, the standardized achievement tests were more skill-specific as two subtests were administered: a reading comprehension test and a listening comprehension test, both of which had been developed by experts and teachers based on the national educational standards and pretests (for a detailed description of the test development procedure, the theoretical rationale of the items used, and their psychometric properties, see Köller et al., 2010; Rupp et al., 2008). In the reading comprehension tests, the students were asked to read a (fictional or non-fictional) text by themselves and to respond to multiple choice and open-ended questions to the text afterwards. To assess listening comprehension, students were asked to listen to a CD with authentic (fictional or non-fictional) texts or conversations and to answer multiple choice and open-ended questions. For all achievement tests, *multiple matrix booklet designs* were used so that each student worked on only a part of the item pool (Gonzalez & Rutkowski, 2010). In both subsamples, the achievement test booklets comprised six item blocks. Three blocks were completed in each of the two 60-minute test sessions. For the first subsample, the achievement test consisted of 28 booklets in total, of which 26 included both English and German blocks whereas two only included German blocks. Within the booklets, item blocks assessing German and item blocks assessing English were alternated. There were a total of 47 different item blocks and 349 single items. On average, each student worked on 102.45 (*SD* =

35.04) items. For the second subsample, the achievement test in French included 24 test booklets. A total of 24 blocks consisting of 403 single items were used with each block appearing in the different booklets at different positions to control for item position effects. On average, each student worked on 39.54 reading items (*SD* = 6.41) and 61.02 listening items (*SD* = 5.60).

In each subsample, the different versions of the test booklet were assigned at random, but students attending the same class worked on the same booklets because of the assessment of listening comprehension. Such planned missingness designs provide an economic and efficient means of collecting data in a limited amount of test time and are thus often applied in large-scale studies (e.g., Enders, 2010; Graham, Hofer, & MacKinnon, 1996; Little, Jorgensen, Lang, & Moore, 2014; Rhemtulla & Little, 2012). All achievement tests were scaled using a one-parameter logistic IRT model (Rasch model). Weighted Likelihood Estimates (WLEs; Warm, 1989) were used to estimate students' ability from the test items using the software ConQuest 2.0 (Wu, Adams, Wilson, & Haldane, 2007). The WLE reliabilities (Wu et al., 2007) were good to excellent for all assessed language skills: German reading: .74; German listening: .80; English reading: .92; English listening: .83; French reading: .84; French listening: .90 (see also Table 1).

Self-concept. Students' self-concepts were assessed by four skill-specific scales and one global scale in the three language domains considered (German, English, French). The same items were used across the three languages and presented in two (related to German and English for Sample 1) respectively three (related to German, English, and French for Sample 2) item blocks. In the instruction of each of these blocks, the students were asked to think of one of the language domains only when answering the items. The four skill-specific scales related to students' self-concepts in reading (e.g., "Reading texts is easy for me"); listening (e.g., "I can easily follow conversations"), writing (e.g., "I am good at writing texts"), and speaking [e.g., "Talking in (German/English/French) is easy for me"] in the respective language domains. The global scale measured students' self-concept in the respective language in a general way not related to any particular language skill [e.g., "I am good at (German/English/French)"; also see Table S1 in the Online Supplements for the item wordings and the descriptive statistics of the items]. All scales (i.e., the four skill-specific and the global scales in the three language domains) consisted of three items to which the students were asked to respond on a 4-point Likert scale ("false", "mostly false", "mostly true", "true"). Higher values on the scales consistently reflect higher levels of self-concept. All scales used in the present study showed adequate to good reliability estimates in terms of

coefficient alpha (α) and additionally in terms of scale reliability (ρ) as an index of reliability explicitly developed in the context of structural equation modelling (Raykov, 2009; see Table 1 for an overview of the used measures and their descriptive statistics and reliability estimates).

Statistical Analyses

Model estimation. We estimated several sets of models within the SEM framework using Mplus 7.11 (Muthén & Muthén, 1998-2013). We first considered the skews and kurtoses of our variables which indicated only slight deviation from normality assumptions (skewness: school grades: range from -0.00 to 0.10; achievement test scores: -0.98 to 0.19; self-concept items: range from -3.33 to 0.25; kurtoses: school grades: range from -0.41 to -0.37; achievement test scores: 0.50 to 5.27; self-concept items: range from -0.91 to 13.45; see Table 1). Subsequently, we selected the robust maximum likelihood (MLR) estimator which has been found to be robust against any violations of normality assumptions and adequate when treating response scales with four categories as continuous variables (e.g., Beauducel & Herzberg, 2006). All models were tested in the three domains of German, English, and French, separately. The first set of models comprises measurement models for testing the multidimensionality of language self-concepts by comparing unidimensional and multidimensional models. In unidimensional models (Models 1-3 in Table 2), the items pertaining to the different verbal skills assessed in this study (reading, writing, speaking, and listening) form one common first-order factor next to a factor for global self-concept defined by its three items (Figure 1a). The multidimensional models (Models 4-6) assume separate first-order factors for each skill-specific self-concept next to the global factor (Figure 1b). Thus, five first-order factors are assumed in these models: Four of them constitute skill-specific self-concept factors (i.e., reading, listening, writing, and speaking self-concepts) each defined by the respective skill-specific items while the global factor represents a general, overarching language self-concept defined by the global self-concept scale. The next series of models (Models 7 to 9) was applied to verify the hierarchy of language self-concepts and each model includes a higher-order factor based on the four skill-specific self-concept factors in addition to the global self-concept factor (Figure 1c).

In the next step, achievement measures were added to the models. Students' scores in the reading and listening tests were included in the 5-factor multidimensional models (Models 4 to 6) leading to Models 10, 11, and 12 in order to probe differential relations between skill-specific self-concept and achievement measures. We then added achievement measures to the higher-order factor models (Models 7 to 9) to examine whether the higher-order factor and the

global factor display differential relations to achievement. First, we used students' school grades as achievement indicators (Models 13 to 15). Given that school grades serve as single item indicators for the achievement factor, their measurement errors were a priori fixed on the basis of the sample variance and an assumed reliability estimate of .95. In a second step, students' achievement test scores were inserted into the higher-order factor models (Models 16 to 18).

In order to test whether boys and girls display a comparable multidimensional and hierarchical structure of language self-concepts, a series of models with certain model parameters increasingly set equal across gender was applied to the higher-order factor models (Models 7 to 9). Following the propositions of Chen, Sousa, and West (2005; see also Cheung, 2008) for invariance testing with higher-order factor models, we started with tests of configural invariance in which only the same factor structure was assumed across groups but all model parameters were freely estimated. In subsequent models, all first-order factor loadings and then all first-order and second-order factor loadings were set to be equal across groups. Invariance of first-order and second-order factor loadings ensures the comparability of the measured constructs across groups and is thus the precondition for all further invariance tests (Millsap, 2011). Subsequently, we conducted models to test whether boys and girls differ in their mean levels of the skill-specific self-concept factors. For this purpose, we fixed the intercepts of the first-order (i.e., skill-specific) factors to zero in one group serving as the reference group (boys in our case) but freely estimated them in the other group (i.e., girls) while constraining the higher-order factor means to zero in both groups (Wang & Wan, 2012). In a next step, we tested for gender differences in the higher-order factors. To this end, the intercepts of the first-order factors were restricted to be zero across groups and the means of the higher-order factors was set to zero in the reference group (i.e., boys) while they were freely estimated in the comparison group (i.e., girls).

Evaluation of model fit. The chi-square statistic can be used for assessing the goodness-of-fit of models conducted in the SEM framework as it represents the discrepancy between the model implied and the observed sample covariance matrices. However, the chi-square value has been demonstrated to be sensitive to sample size as even a negligible discrepancy between the model implied and observed sample covariance matrices is likely to result in a significant chi-square statistic leading to model rejection (Marsh, Hau, & Grayson, 2005). Hence, researchers are advised to rather rely on descriptive goodness-of-fit indices. We thus present the most commonly used goodness-of-fit indices such as the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation

(RMSEA), and the standardized root mean square residual (SRMR). For the CFI, values above .90 were originally deemed indicative of an adequate model fit (Bentler, 1990) although more recent guidelines (Hu & Bentler, 1998, 1999) recommend the stricter cut-off value of .95 as a criterion for good model fit. With respect to the TLI, .95 is also recommended as the cut-off value above which good model fit can be claimed (Hu & Bentler, 1998, 1999). Browne and Cudeck (1993) introduced guidelines for interpreting RMSEA values according to which values about .05 indicate "close fit", values near .08 indicate "fair fit", and values above .10 indicate "poor fit". Regarding the SRMR, values between .08 and .10 are interpreted as a good model fit (e.g., Kline, 2005) but Hu and Bentler (1999) only consider values close to .08 as a good model fit. Given the controversial cut-off values for the various goodness-of-fit indices, the interpretation of model fit always remains partially subjective and researchers are thus advised to simultaneously consider the results of various fit indices along with parameter estimates, statistical conformity, and theoretical adequacy of the model (see Marsh, Hau, & Wen, 2004).

Our research questions involve the inspection of nested models. Models are nested within each other when one model is a specific subtype of a more general model. Hence, certain model parameters are fixed in the nested model while freely estimated in the more general model. The comparison of nested models is used in this study to explore whether two factor correlations, for example the correlations between skill-specific self-concept and achievement measures, are of the same or different size. In this context, a more restrictive model assuming equal-sized correlations is compared to a less restrictive model with freely estimated factor correlations. Nested models also occur in the context of multi-group invariance testing (gender in this study) as the respective tests comprise a series of models which only differ in the model parameters set to be equal across groups. Because of the high sensitivity of the chi-square value to sample size, researchers are advised to inspect different goodness-of-fit indices for the evaluation and comparison of nested models (Marsh, Hau, & Grayson, 2005). According to Cheung and Rensvold (2002; see also Chen, 2007), invariance can be assumed as long as the CFI does not decrease more than .01 and as long as the RMSEA does not increase by more than .015 between less and more restrictive models. These recommendations, however, have been considered too liberal by some researchers (e.g., Meade, Johnson, & Braddy, 2008). The guidelines of Cheung and Rensvold should also be reconsidered with respect to their applicability to mean level differences (Chen et al., 2008; Fan & Sivo, 2009). Thus, when testing gender differences in factor means, we additionally take effect sizes into account. In sum, with respect to both the simple assessment of model fit

and the evaluation of nested models, researchers are advised to use proposed cut-off values only as rough guidelines. They should rather additionally consider the resulting models including their parameters in conjunction with substantive theory, and take all further information into account (Marsh et al., 2004).

Treatment of missing data. Due to the multiple matrix sampling method applied to the achievement tests, not all students of the first subsample received booklets that included test items related to all specific verbal skills. This resulted in a considerable amount of missing data on the language-specific achievement test scores in the first subsample. Achievement test scores in English reading and English listening were missing for 401 students (14.9%) who only worked on German language tests. Achievement test scores in German reading were missing for 788 students (28.7%); achievement test scores in German listening were missing for 2006 students (72.9%). However, these missing data, which were planned, can be considered to be missing completely at random (MCAR; Schafer & Graham, 2002). There were no missing data on the achievement test scores in the second subsample as all students worked on both listening and reading tests in French.

With regard to the self-concept items, the amount of missing data was very small and ranged from 0.5% to 2.3% in the first subsample and from 0.4% to 2.7% in the second subsample. Full Information robust Maximum Likelihood (FIML) estimation was used to handle missing data which is known to be a trustworthy and efficient procedure that yields unbiased estimates under the missing at random assumption while retaining power (Enders, 2010; Graham, 2009).

Results

Within-Structure of Language Self-concepts

The 5-factor models (Models 4 to 6 in Table 2; Figure 1b) which assumed separate factors for the different skill-specific self-concepts (i.e., reading, listening, writing, and speaking self-concepts) besides a global self-concept resulted in adequate to good levels of fit to the data. By contrast, the models which stated a unidimensional factor for all skill-specific measures besides a global self-concept factor (Models 1 to 3; Figure 1a) did not show adequate fits to the data. It is interesting to note that on the basis of the CFI and TLI values, the 5-factor model was found to more adequately represent the structure of students' self-concepts related to foreign languages (English, French) than to students' native language (German). The self-concept facets were all well defined in all three languages as it is evident from their substantial and positive standardized factor loadings (Figure 1b). Table 3 depicts the standardized correlations between the self-concept factors resulting from the 5-factor

models. In all languages considered, the correlations among the four skill-specific self-concept factors were substantial allowing for tests of higher-order factor models. These correlations were, however, far from perfect providing evidence for the students' ability to differentiate between self-concepts related to different skills indicating multidimensionality within language self-concepts. In this regard, it is worthwhile to point out that the different skill-specific self-concept factors were found to be on average more interrelated within the domains of English (range from $r = .68$ to $r = .87$; mean $r = .79$) and French (range from $r = .74$ to $r = .90$; mean $r = .85$) as the two foreign languages compared to the domain of German as students' native language (range from $r = .53$ to $r = .76$; mean $r = .64$). A particularly high correlation emerged between the self-concepts for speaking and listening in French ($r = .90$).

The higher-order models demonstrated adequate to good levels of fit (Models 7 to 9). In all models, the higher-order factor was well defined by substantial loadings of the first order skill-specific self-concept factors (see Figure 1c). The correlation between the higher-order and global factors was high in each of the three languages considered, with slightly higher relations in English ($r = .95$, $p < .001$) and French ($r = .95$, $p < .001$) compared to German ($r = .85$, $p < .001$). These findings support the assumption of the coexistence of multidimensionality and hierarchy within specific language self-concepts as the results demonstrate the existence of an overarching general language self-concept incorporating skill-specific facets. This pattern of findings was demonstrated to be similar across the three languages considered.

Relations to Achievement

To further substantiate the notion of multidimensionality, we considered the relations between the various skill-specific self-concept facets and students' achievement in the reading and listening achievement tests (Table 4). In English (Model 11) and French (Model 12), students' reading test scores were found to be similarly related to each of the five self-concept factors (English: $r = .43$ to .47; French: $r = .48$ to .52; Table 4). Students' scores in the listening achievement test also showed similar relations to each of the five self-concept factors (English: $r = .38$ to .45; French: $r = .47$ to .57). Thus, contrary to the assumption of skill-specific self-concept–achievement relations, reading self-concept did not show higher relations to the reading test score and listening self-concept did not demonstrate higher relations to the listening test score compared to the other skill-specific self-concept facets. For German (Model 10), reading self-concept descriptively showed higher relations to the reading test score ($r = .45$) compared to the other self-concept facets ($r = .30$ to .35). However, a model assuming equal-sized relations among the reading test score and all self-concept facets

did not show a decline in model fit [CFI = .930 (Δ = -.007), TLI = .908 (Δ = -.007), RMSEA = .060 (Δ = +.002), SRMR = .061 (Δ +.020)]. Hence, although the descriptive results for German indicate the expected pattern of findings supporting a skill-specific multidimensional self-concept structure due to the relatively highest relation between reading self-concept and reading achievement, this finding could not be retained in significance tests.

In the subsequent models (Models 13-15), we examined and compared the relations of school grades to the higher-order and global self-concept factors. Table 5 indicates that, irrespective of the language considered, school grades showed slightly higher relations to the global factor than to the higher-order factor. However, constraining the relations of school grades to the higher-order and global factors to be equal (Models 13a-15a) led to only marginal declines in model fits compared to the models with freely estimated achievement relations. A closer examination reveals that the model fit declined to a greater extent when assuming invariant relations of the higher-order and global factors to school grades in German (Model 13a: ΔCFI = -.09; ΔTLI = -.10) relative to the results emanating from the models for English (Model 14a: Δ CFI = -.02; ΔTLI = -.02) and French (Model 15a: ΔCFI = -.05; ΔTLI = -.06). In fact, for this constrained model for German self-concept, the TLI even dropped to a range implicating poor model fit. This finding suggests that the difference in the correlations of the higher-order and global factors to school grades is more substantial in German than in English and French.

The next series of models (Models 16 to 18) investigated the relations of verbal achievement test scores to the higher-order and global self-concept factors. Given that the reading and listening test scores did not demonstrate differential relations to the skill-specific first-order self-concept facets and in light of the substantial correlations between the reading and listening test scores (German: $r = .59$, English: $r = .75$, French: $r = .73$; all $p < .001$), here, we operationalized achievement by a common latent factor defined by both the reading and listening test scores in each language. The descriptive pattern was reversed to that found for school grades. Across the three languages, the higher-order factors consistently demonstrated slightly higher relations to the achievement test scores than the global factor. However, the results of the restrictive models (Models 16a, 17a, 18a), where the relations of the achievement test scores were constrained to be of equal size for the higher-order factor and the global factor, displayed only marginal changes in model fit. These findings thus argue against differential achievement relations for the higher-order and global factors. In other words, since the higher-order factor and global factor were found to be similarly related to

achievement test scores in all three languages (German, English, and French), they can be taken to be similar constructs.

Gender Differences

Gender invariance was first tested by examining whether boys and girls displayed a similar or differential structure of multidimensional and hierarchical language self-concepts. We therefore tested models of configural invariance of the higher-order models assuming that boys and girls hold the same structure of language self-concepts (Model I1 for German, Model I6 for English, Model I11 for French, Table 6). These models resulted in adequate levels of fit so that boys and girls were found to similarly display multidimensional and hierarchical structures within German, English, and French self-concepts. The results further revealed equal-sized first-order (Model I2 for German, Model I7 for English, Model I12 for French) and second-order factor loadings (Model I3 for German, Model I8 for English, Model I13 for French) across groups given the small decreases in the descriptive fit indices relative to the models of configural invariance. These findings attest that the same constructs associated with the same meanings are measured across gender groups and thus constitute the preconditions for further invariance tests.

In a next step, we tested whether the various skill-specific self-concept facets are subject to similar or differential gender effects. For this purpose, the intercepts of the first-order factors (reading, writing, speaking, listening, and global) were set to zero in the group of boys serving as the reference group and freely estimated for girls while constraining the second-order factor means to zero in both groups (Model I4 for German, Model I9 for English, Model I14 for French). In this case, girls' mean levels were found to significantly positively deviate from those of boys for all first-order factors except for the speaking factors in German (in *SD* units: .014, $p = .071$) and English (in *SD* units: .021, $p = .325$; see Figure 2 and Table 1). Thus, girls were found to display higher mean levels on all skill-specific self-concept factors and on the global self-concept factors referring to German, English, and French with the exception that girls and boys showed similar-sized mean levels in German and English speaking self-concepts.

We then examined whether the higher-order and global factors display similar or differential gender effects (Model I5 for German, Model I10 for English, Model I15 for French). We therefore constrained the first-order factor intercepts to zero in both groups. As the higher-order factor means were fixed to zero in the first group as the reference group (i.e., boys) by default for identification purposes, girls' freely estimated mean levels depicted the deviation from boys' mean levels. Girls were found to demonstrate significantly higher levels

of higher-order factors in all three languages (in *SD* units: German: $d = .305$; English: $d = .128$; French: $d = .332$; all $p < .001$). Along with findings from the aforementioned models (Model I4 for German, Model I9 for English, Model I14 for French) revealing higher mean levels for girls on the global self-concept factors, we could suggest that girls displayed higher mean levels on both the higher-order and global factors in all three languages of German, English, and French.

Discussion

Research on verbal self-concept so far has focused predominantly on its separability from math self-concept (Marsh, 1986, 1990; Möller et al., 2009) and its differentiation into self-concepts for specific languages (Marsh et al., 2001; Marsh & Yeung, 2001; Möller et al., 2006; Xu et al., 2013; Yeung & Wong, 2004), and therefore places a strong focus on its multidimensionality. With respect to the structure within self-concepts related to specific languages, Lau et al. (1999) and Yeung et al. (2000) provided first evidence of the simultaneous operation of multidimensionality and hierarchy. The present study aimed to extend this evidence by considering achievement relations and gender differences. In addition, we aimed to test whether the extent of multidimensionality and hierarchy varies across languages by including students' native language and two foreign languages.

Overall, our results supported the notion of a simultaneous operation of hierarchy and multidimensionality within the structures of self-concepts related to German as students' native language and English and French as students' foreign languages. Evidence of a multidimensional structure within language self-concepts was provided since the models distinguishing among skill-specific self-concepts in reading, listening, writing, and speaking fitted the data considerably better than unidimensional models. In addition, the skill-specific self-concept facets were found to be highly but not perfectly related to each other further supporting their distinctiveness. Evidence of a hierarchical structure within language self-concepts was provided as the skill-specific self-concepts related to reading, listening, writing, and speaking could be combined into a superordinate higher-order factor which was found to be highly correlated with an independently assessed global self-concept factor. Our findings thus replicate the insights obtained from Lau et al. (1999) and Yeung et al. (2000) in terms of the coexistence of multidimensional and hierarchical language self-concepts. This notion was further substantiated but also qualified by the considerations of achievement relations and gender differences as additional analytical approaches to test the structure within language self-concepts. In addition, although the findings seem to be similar across the three languages considered in this study, subtle, but interesting differences emerged.

Further Support for Multidimensionality

Support for the multidimensional nature of language self-concepts originates from the inspection of gender differences. Consistent with previous research (e.g., Eccles et al., 1993; Jacobs et al., 2002; Marsh, 1989; Watt, 2004; Wilgenbusch & Merrell, 1999), girls were found to show higher levels of global language self-concepts and reading self-concepts across the three languages considered. The girls' superiority was also demonstrated for writing and listening self-concepts in all three languages. However, boys and girls were not found to differ in their self-concepts of speaking in German and English. Thus, the consideration of only global language self-concepts would indicate higher self-concept levels for girls, but would neglect the fact that boys measure up at least in the specific skill of speaking. The strongest gender effects were observed for skill-specific and global self-concepts related to French with girls demonstrating higher levels than boys. This result is consistent with reports that French is perceived as a feminine language leading to a large gender gap in student motivation for learning French (Williams, Burden, & Lanvers, 2002).

The findings resulting from the investigation of relations between skill-specific language self-concepts and test scores for reading and listening achievement, however, do not support a multidimensional structure. In all languages examined, reading achievement was not found to be more strongly related to reading self-concept compared to the other skill-specific self-concept facets. In parallel, listening achievement did not demonstrate relatively higher relations to listening self-concept.

Therefore, on the one hand, the findings on the structure and gender differences related to specific language self-concepts underpin their multidimensional nature. On the other hand, this notion could not be supported by the relations between skill-specific self-concepts and achievement test scores. This finding might be regarded as evidence of a weak multidimensional structure but in favor of the existence of a unitary language self-concept. Although it would indeed be reasonable to draw such a conclusion, particularly in light of the findings supporting hierarchy within language self-concepts outlined below, caution should be exercised regarding the exclusion of multidimensionality within language self-concepts. The high correlation between reading and listening achievement tests as well as the high correlation among the various skill-specific self-concept facets might have precluded a pattern of skill-specific self-concept–achievement relations. Furthermore, the present study only included achievement tests with regard to the specific skills of reading and listening but did not incorporate achievement tests for speaking and writing. Thus, there were no matching achievement tests for each skill-specific self-concept assessed. Consequently, future research

is necessary to investigate the skill-specific nature of language self-concepts by considering the relations between skill-specific self-concepts and outcome measures. In this context, it might also be worthwhile to consider other outcome criteria such as motivation or behavior.

Finally, these results match previous findings and conceptualizations of the structure of academic self-concept. Accordingly, self-concept facets are harder to distinguish at finer-grained levels of specificity. For instance, math and verbal self-concepts present two self-concepts that are clearly related to different domains and have been consistently found to be weakly correlated (see for example the average correlation of $r = .10$ in the meta-analyses of Möller et al., 2009). When, however, only considering the verbal domain, the correlation between self-concepts for different languages gets higher (see for example the correlation of $r = .45$ between English and Chinese self-concepts in Xu et al., 2013). Self-concept correlations rises even higher within specific languages (i.e., between skill-specific self-concepts for separate languages as examined in this study) indicating a less pronounced multidimensionality. Therefore, it remains questionable whether a further differentiation within skill-specific self-concepts (e.g., different aspects of writing self-concepts) is theoretically and practically reasonable (but see Henschel, Roick, Brunner, & Stanat, 2013 for evidence of a further differentiation within reading self-concept dependent upon text type, i.e., non-fictional vs. literary reading self-concepts).

Further Support for Hierarchy

Hierarchy within language self-concepts was first supported by the appropriateness of higher-order factor models and the substantial correlations between the higher-order factors and the global factors. This finding implicates that both constructs are similar to each other and that the global factor is an adequate representation of the various skill-specific self-concepts. This conclusion was further supported by the finding of substantial correlations between the higher-order factors and achievement (both in terms of achievement test scores and school grades) since these relations were similar to the relations demonstrated for the global factors. In addition, girls displayed higher mean levels on both the higher-order and global factors further arguing for the similarity of both constructs and for the existence of hierarchy within specific language self-concepts.

Despite these supportive findings, the notion of hierarchy within language self-concepts has to be qualified. The evaluation of the invariance tests applied to examine whether the higher-order and global factors display differential or equivalent relations to achievement (i.e., achievement test scores and school grades) is based on changes in the descriptive goodness-of-fit indices. However, a closer inspection of the descriptive results

suggests a more differentiated picture as the higher-order factors were more highly related to achievement test scores while the global factors were more highly related to school grades. This observation indicates that the higher-order and global factors might be distinct due to differential relations to outcome criteria. This should be further examined beyond the present study, for example when taking other outcome variables than achievement into account.

In sum, neither the findings supporting multidimensionality nor the findings supporting hierarchy were without qualifications, therefore strengthening the assumption that both structural features (i.e., multidimensionality and hierarchy) exist within language self-concepts. These findings are valuable since self-concept research has made a strong case for the multidimensionality rather than the hierarchy of verbal self-concept given the differentiation between math and verbal self-concepts (Marsh, 1986, 1990; Möller et al., 2009) and between self-concepts for various languages (Marsh et al., 2001; Marsh & Yeung, 2001; Möller et al., 2006; Xu et al., 2013; Yeung & Wong, 2004). Taking a within-perspective and when considering self-concepts referring to specific languages, the situation seems different, and multidimensionality and hierarchy seem to coexist.

Generalizability across Languages

At a first glance, the findings of a multidimensional and hierarchical self-concept structure seemed to be fairly consistent across the three languages considered (German, English, and French). However, a closer look reveals some differences between the findings for German as students' native language and English and French as foreign languages. The skill-specific self-concept facets related to German were found to be on average less interrelated than those for English and French implying that students might have a more differentiated self-concept for German. These results match previous findings revealing a more unified and less differentiated structure of skills in foreign languages (e.g., Leucht, Retelsdorf, Möller, & Köller, 2010). For instance, analyzing the factor structure underlying the TOEFL, Jang and Roussos (2007) found a combined factor for the two subtests Listening Comprehension and Structure and Writing Expression. The notion that students differentiate less between verbal skills in foreign than in native languages is further supported by the descriptive inspection of the correlations between the skill-specific self-concepts and achievement measures. For English and French, there was no indication of skill-specific relations between self-concept and achievement measures, as for example reading and listening test scores demonstrated similar relations to reading self-concept. For German, however, the resulting pattern of relations was more indicative of a multidimensional, skill-specific self-concept as reading self-concept was more highly related to reading achievement.

While a closer inspection of our findings thus suggests a stronger manifestation of multidimensionality within the native language German self-concept, the results in turn imply a weaker hierarchy. The higher-order and global factors in German tend to differ from each other to a larger extent than in English and French. For English and French, the higher-order and global factors were found to be similarly related to school grades and achievement test scores. For German, in contrast, the pattern of self-concept–achievement relations seemed to vary contingent upon the achievement indicator. Whereas the higher-order and global factors were similarly related to school grades, the higher-order factor was found to display higher relations to achievement test scores than to school grades. The test scores exhibited a more skill-specific perspective as they involved specific tests for reading and listening achievement. Self-concept facets related to these two skills are integrated into the higher-order factor of German self-concept which might have led to its higher relations to test scores. In other words, the stronger multidimensional, skill-specific nature of German self-concept might be reflected in higher relations between the higher-order factor integrating these skill-specific self-concept factors and a skill-specific achievement outcome (i.e., achievement test scores here) than between the higher-order factor and broader, skill-spanning achievement indicators (i.e., school grades).

These findings could be explained by the fact that German is students' native language while English and French are foreign languages for the students analyzed here. Foreign languages are, at least for the majority of the students, mostly acquired, practiced, and experienced at school. The teaching of different foreign language skills takes place simultaneously in foreign language classes, and students get a single grade covering their achievement in the different skills. By contrast, as their native language, students consistently apply their German skills in a broader scope of school subjects and life domains leading to more opportunities for the students to evaluate and compare their individual skill performances and to receive differential skill-specific feedback possibly resulting in more differentiated skill-specific self-concepts.

The findings of the present study which hence suggest some differences in the within-structures of native language (German) and foreign language (English, French) self-concepts, however, have to be interpreted carefully. First of all, the differences were small and the interpretations were based on descriptive observations. Moreover, the two subsamples of the present study differed as the students from the second subsample mostly attended the academic secondary school track and the proportion of girls was larger. Even though it was thus difficult to compare the groups of students with English and French as foreign languages,

the main conclusions regarding presumptive differences between native (German) and foreign (English, French) languages remain largely unaffected since the students in both samples replied to the measures for German language. This issue, however, highlights the importance of including students' social, cultural, and educational environment as a further determinant of self-concept. Foreign language self-concepts might be particularly sensitive to contextual issues since students' foreign language motivation, as a key construct in foreign language acquisition in general, has also been found to vary with students' context (e.g., Csizér, & Dörnyei, 2005; Dörnyei, 2001; Gardner, 2010; Masgoret & Gardner, 2003). Further studies should therefore examine the generalizability of the findings across students of different ages, educational systems, countries, and ability tracks.

Limitations, Future Research, and Practical Implications

Self-concept research and theory might benefit from this further empirical support for the simultaneous operation of multidimensionality and hierarchy within language self-concepts. Irrespective of the interesting findings of the present study, future research is certainly necessary to further explore the within-structure of language self-concepts. The present study is only cross-sectional in nature and focuses on ninth-grade German students. Longitudinal studies might be worthwhile to gain insight into the development of the multidimensional and hierarchical structure. Self-concept has been found to become increasingly differentiated with students' age (Marsh & Ayotte, 2003) so that it might be worthwhile to compare the structure within language self-concepts between younger and older students. Furthermore, longitudinal studies enable the examination of the interrelation among factors across time, as it might be interesting to investigate whether the global factor emerges first, subsequently affecting the multiple skill-specific self-concepts, or conversely (cf., top-down or bottom-up effects within self-concept; Marsh & Yeung, 1998b). The cross-sectional design also restricts the examination of self-concept–achievement relations, which have been found to be reciprocal in nature (Marsh & Craven, 2006).

In the light of the strong focus of self-concept research and theory on the multidimensionality and domain specificity of self-concept but given the present findings of the co-existence between multidimensional and hierarchy, the hierarchy of self-concept might re-enter the agenda (see for example Brunner et al., 2010). The present study has been restricted to higher-order factor models to depict the hierarchical structure of self-concept. Future studies, therefore, should consider alternative models such as bifactor models (Brunner et al., 2010; Brunner, Nagy, & Wilhelm, 2012) which might be a viable approach to take into account both the multidimensionality and hierarchy of skill-specific verbal self-concepts.

Bifactor models propose a global self-concept that underlies all answers to items purported to reflect various domain-specific self-concept facets and that co-exists with specific factors reflecting the part of the items unexplained by the global factor (Holzinger & Swineford, 1937).

Although our findings first suggest the generalizability of a multidimensional and hierarchical structure across the self-concepts related to students' native language and two different foreign languages, a closer look revealed some differences between the structures of native and foreign language self-concepts. This observation might be a good starting point to further studies on similarities and differences of native and foreign language self-concepts. Besides, our findings might inspire researchers to investigate similarities and differences of the self-concepts for different foreign languages as we did not expand on the comparison between English and French self-concepts. Foreign languages, however, vary contingent upon their cognitive demands, students' learning experiences, and instructional approaches which might lead to differences in the respective self-concepts including their internal structures, developmental trajectories, and outcome relations.

Going even further beyond the verbal domain, it might be interesting to study the applicability of the presented insights to other domains such as math. For example, the potentially multidimensional and hierarchical structure of self-concepts related to different specific skills in math such as algebra and geometry could be examined. This would also allow for elucidating how far multidimensionality goes within specific self-concepts and to what extent this is theoretically and practically relevant.

Besides further research on the construct of language self-concepts, it might also be worthwhile to consider the within-structure of other related motivational constructs. In this context, for instance, research might be interested in examining the domain-related and skill-related differentiation of goal orientations, value beliefs, or self-efficacy (Bong, 2001; Marsh et al., 2013). The construct of self-efficacy is closely related to that of self-concept given that they both relate to students' self-perceptions of competence (Bong & Skaalvik, 2003). However, both constructs can be separated theoretically (Bong & Skaalvik, 2003) and empirically (e.g., Ferla, Valcke, & Cai, 2009; Jansen, Scherer, & Schroeders, 2015; Parker et al., 2014). Self-concept refers to students' self-perceptions of competence related to a given domain in general (e.g., a school subject or a sub-facet of a school subject) and is past-orientated. Self-efficacy, in contrast, refers to an individual's expectation and confidence to successfully carry out a specific behavior (e.g., to solve a specific task) and is thus future-orientated. This can be illustrated by an item example used by Pajares, Miller, and Johnson

(1999) to measure writing self-efficacy which reads: “How sure are you that you can correctly spell all words in a one-page story or composition?” Similarly, Shell, Colvin, and Bruning (1995) asks for students’ subjective probability to successfully complete tasks such as “know all the words on a page in one of your school books“ to measure reading self-efficacy (but see Bong & Skaalvik, 2003 for a discussion and illustration of inconsistencies in the assessment of self-efficacy). As can be seen from these items also relating to verbal skills, they are conceptually different from the self-concept items used in our study. However, it should be mentioned that, given the conceptual definitions of both constructs, it seems likely that self-concept and self-efficacy measures become harder to distinguish when the addressed content domain becomes more specific (e.g. writing self-concept and writing self-efficacy may be harder to differentiate than English self-concept and English self-efficacy). The task-orientation as an important part of the conceptualization of self-efficacy is partly confounded with the specificity level of items (including self-concept items) because more specific items are more likely to include references to concrete tasks. Thus, the empirical differentiation of self-concept and self-efficacy on the level of specific verbal skills would be interesting to examine in further research.

Besides research on the differentiation of self-concept and self-efficacy, future research seems also to be necessary on the discriminability between various self-efficacy beliefs. So far, research on the internal structure of self-efficacy has demonstrated its multidimensionality in terms of domain specificity, which has, however, found to be less pronounced compared to the domain specificity of academic self-concept (Bong, 2001; Bong & Hocevar, 2002; Skaalvik & Rankin, 1995). Given that these studies so far have only addressed self-efficacy judgement between domains (e.g., math, verbal), more research would thus be interesting to study the differentiation of self-efficacy beliefs within domains. A less pronounced multidimensionality might implicate strong hierarchy, which, however, does not seem to be true since Bong and Skaalvik (2003) ascribe self-efficacy a ‘loosely hierarchical’ structure.

Findings from this study might also be beneficial for educators. The finding that hierarchy seems to exist within specific language self-concepts might facilitate the application of global self-concept measures, i.e., the assessment of students’ overall, global self-perceptions related to specific languages. This parsimonious approach is implemented in recognized self-concept instruments such as the Self-Description Questionnaires (SDQ, Byrne, 1996; Marsh, 2007) and seems to give a warrantable estimation of students’ self-concept across the range of skill-specific self-concepts. Yet, practitioners should consider the

limitations of global scales such as the insufficient comprehension of gender differences (see our findings on equivalent mean levels for boys and girls in speaking self-concepts for German and English). Therefore, in some instances, such a global approach may be too superficial and a more fine-grained approach could provide practitioners or researchers with more specific information on students' self-perceived skills which may help disclose their self-perceived strengths and weaknesses. These pieces of information could then be useful for intervention strategies which should be specifically tailored to address target domains of self-concepts in order to be most effective (O'Mara, Marsh, Craven, & Debus, 2006).

References

Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology, 102*, 281–298. doi: 10.1037/a0019318.

Arens, A. K., Yeung, A. S., Craven, R. G., & Hasselhorn, M. (2013). A short German version of the Self-Description Questionnaire I: Theoretical and empirical comparability. *International Journal of Research & Method in Education, 36*, 415–438. doi:10.1080/1743727X.2012.710503

Arens, A. K., Yeung, A. S., & Hasselhorn, M. (2014). Native language self-concept and reading self-concept: Same or different? *The Journal of Experimental Education, 82*, 229–252. doi: 10.1080/00220973.2013.813362

Aunola, K., Leskinen, E., Onatsu-Arvilommi, T., & Nurmi, J-E. (2002). Three methods for studying developmental change: A case of reading skills and self-concept. *British Journal of Educational Psychology, 72*, 343–364.doi: 10.1348/000709902320634447

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

Bentler, P. M. (1990). Comparative fit indexes in structural models. P*sychological Bulletin, 107*, 238–246. doi: 10.1037/0033-2909.107.2.238

Berninger, V. (2000). Development of language by hand and its connections to language by ear, mouth, and eye. *Topics in Language Disorders, 20*, 65–84.

Bong, M. (2001). Between- and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task value, and achievement goals. *Journal of Educational Psychology, 93,* 23–34. doi: 10.1037/0022-0663.93.1.23

Bong, M., & Hocevar, D. (2002). Measuring self-efficacy: Multitrait-multimethod comparison of scaling procedures. *Applied Measurement in Education, 15*, 143–171. doi: 10.1207/S15324818AME1502_02

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*, 1–40. doi: 10.1023/A:1021302408382

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, CA: Sage.

Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology, 102*, 964–981. doi: 10.1037/a0019644

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80*, 796–846. doi: 10.1111/j.1467-6494.2011.00749.x

Buck, G. (2001). *Assessing Listening*. Cambridge University Press: New York.

Byrne, B. M. (1996). *Measuring self-concept across the life span: Issues and instrumentation*. Washington, DC: American Psychological Association.

Chapman, J. W., & Tunmer, W. E. (1995). Development of young children's reading self-concepts: An examination of emerging subcomponents and their relationship with reading achievement. *Journal of Educational Psychology, 87*, 154–167. doi: 10.1037/0022-0663.87.1.154

Chen, F. F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504. doi: 10.1080/10705510701301834

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12*, 471–492. doi: 10.1207/s15328007sem1203_7

Chen, S.-K., Yeh, Y.-C., Hwang, F.-M., & Lin, S. S. J. (2013). The relationship between academic self-concept and achievement: A multicohort–multioccasion study. *Learning and Individual Differences, 23,*172–178. doi: 10.1016/j.lindif.2012.07.021

Cheung, G. W. (2008). Testing equivalence in the structure, means, and variances of higher-order constructs with structural equation modeling. *Organizational Research Methods, 11*, 593-613. doi: 10.1177/1094428106298973

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

Csizér, K., & Dörnyei, Z. (2005). Language learners‘ motivational profiles and their motivated learning behavior. *Language Learning, 55*, 613-659.

Dörnyei, Z. (2001). New themes and approaches in second language motivation research. *Annual Review of Applied Linguistics, 21*, 43–59.

Eccles, J. S., Wigfield, A., Harold, R., & Blumenfeld, P. B. (1993). Age and gender differences in children’s self- and task perceptions during elementary school. *Child Development, 64*, 830–847. doi: 10.1111/j.1467-8624.1993.tb02946.x

Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders, 20* (3), 19–36.doi: 10.1097/00011363-200020030-00005

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

Fan, X., & Sivo, S. A. (2009). Using goodness-of-fit indexes in assessing mean structure invariance. *Structural Equation Modeling, 16*, 54–69. doi: 10.1080/10705510802561311

Ferla, J., Valcke, M., & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, *19*, 499–505. doi: 10.1016/j.lindif.2009.05.004

Gardner, R. C. (2010). Motivation and second language acquisition. The socio-educational model. New York: Peter Lang.

Gonzalez, E., & Rutkowski, L. (2010). Principles of multiple matrix booklet designs and parameter recovery in large-scale assessments. *IERI Monograph Series*, *3*, 125–156.

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

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, *31*, 197–218. doi: 10.1207/s15327906mbr3102_3

Harris, D. P. (1969). *Testing English as a second language*. New York: McGraw-Hill.

Hattie, J. (2008). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: New York: Routledge.

Henschel, S., Roick, T., Brunner, M. & Stanat, P. (2013). Leseselbstkonzept und Textart: Lassen sich literarisches und faktuales Leseselbstkonzept trennen [Reading self-

concept and text type: Can literary and factual reading self-concepts be differentiated]? *Zeitschrift für Pädagogische Psychologie, 27*, 181–191. doi: 10.1024/1010-0652/a000103

Holzinger, K. J., & Swineford, F. (1937). The bi-factor model. *Psychometrika, 2*, 41–54. doi: 10.1007/BF02287965

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424–453. doi: 10.1037/1082-989X.3.4.424

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

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, 49, 505–528. doi: 10.1016/j.jsp.2011.07.001

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grade one through twelve. *Child Development, 73*, 509–527. doi: 10.1111/1467-8624.00421

Jang, E. E., & Roussos, L. (2007). An investigation into the dimensionality of TOEFL using conditional covariance-based nonparametric approach. *Journal of Educational Measurement, 44*, 1–21. doi: 10.1111/j.1745-3984.2007.00024.x

Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, *41*, 13–24. doi: 10.1016/j.cedpsych.2014.11.002

Katzir, T., Lesaux, N. K., & Kim, Y-S. (2009). The role of reading self-concept and home literacy practices in fourth grade reading comprehension. *Reading and Writing, 22*, 261–276. doi: 10.1007/s11145-007-9112-8

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guildford.

Köller, O., Knigge, M., & Tesch, B. (2010). *Sprachliche Kompetenzen im Ländervergleich.* [Comparing German federal states with regard to verbal abilities] Münster: Waxmann.

Lado, R. (1961). *Language testing: The construction and the use of foreign language tests*. London: Longman.

Lau, I. C-Y., Yeung, A. S., Jin, P., & Low, R. (1999). Toward a hierarchical, multidimensional English self-concept. *Journal of Educational Psychology, 91*, 747–755. doi: 10.1037/0022-0663.91.4.747

Leucht, M., Retelsdorf, J., Möller, J., & Köller, O. (2010). Zur Dimensionalität rezeptiver fremdsprachlicher Kompetenzen. [On the dimensionality of receptive skills in English as a foreign language]. *Zeitschrift für Pädagogische Psychologie, 24*, 123–138. doi: 10.1024/1010-0652/a000010

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2014). On the joys of missing data. *Journal of Pediatric Psychology*, *39*, 151–162. doi: 10.1093/jpepsy/jst048

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23*, 129–149. doi: 10.3102/00028312023001129

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology, 82*, 417–430. doi: 10.1037/0022-0663.81.3.417

Marsh, H. W. (1990). Influences of internal and external frames of reference on the formation of math and English self-concepts. *Journal of Educational Psychology, 82*, 107–116. doi: 10.1037/0022-0663.82.1.107

Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology.* Leicester, UK: British Psychological Society.

Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., Xu, M. K., Nagengast, B., & Parker, P. (2013). Factor structure, discriminant and convergent validity of TIMSS math and science motivation measures: A comparison of USA and Saudi Arabia. *Journal of Educational Psychology, 105*, 108–128. doi: 10.1037/a0029907

Marsh, H. W., & Ayotte, V. (2003). Do multiple dimensions of self-concept become more differentiated with age? The differential distinctiveness hypothesis. *Journal of Educational Psychology, 95*, 687–706. doi: 10.1037/0022-0663.95.4.687

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*, 133–163. doi: 10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary psychometrics* (pp. 275-340). Mahwah NJ: Erlbaum.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler's (1999). *Structural Equation Modeling, 11*, 320–341. doi: 10.1207/s15328007sem1103_2

Marsh, H. W., Kong, C. K., & Hau, K. T. (2001). Extension of the internal/external frame of reference model of self-concept formation: Importance of native and nonnative languages for Chinese students. *Journal of Educational Psychology, 93,* 543–553. doi: 10.1037/0022-0663.93.3.543

Marsh, H. W., Kuyper, H., Seaton, M., Parker, P. D., Morin, A. J. S., Möller, J., & Abduljabbar, A. S. (2014). Dimensional comparison theory: An extension of the internal/external frame of reference effect on academic self-concept formation, *Contemporary Educational Psychology, 39*, 326-341. doi:10.1016/j.cedpsych.2014.08.003

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology, 8*, 59–77. doi: 10.1348/000709910X503501

Marsh, H. W. & O'Mara, A. J. (2008). Self-concept is as multidisciplinary as it is multidimensional. In H. W. Marsh, R. G. Craven & D. M. McInerney (Eds.), *Self-processes, learning, and enabling human potential. Dynamic new approaches* (pp. 87–115). Charlotte, NC: Information Age.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades and standardized test scores: Reciprocal effects model of causal ordering. *Child Development, 76*, 397–416. doi: 10.1111/j.1467-8624.2005.00853.x

Marsh, H. W., & Yeung, A. S. (1997). Coursework selection: The effects of academic self-concept and achievement. *American Educational Research Journal, 34*, 691–720. doi: 10.3102/00028312034004691

Marsh, H. W., & Yeung, A. S. (1998a). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal, 35*, 705–738. doi:10.3102/00028312035004705

Marsh, H. W., & Yeung, A. S. (1998b). Top-down, bottom-up, and horizontal models: the direction of causality in multidimensional, hierarchical self-concept models. *Journal of Personality and Social Psychology, 75*, 509–527. doi: 10.1037/0022-3514.75.2.509

Marsh, H. W., & Yeung, A. S. (2001). An extension of the internal/external frame of reference model: A response to Bong (1998). *Multivariate Behavioral Research, 36,* 389–420. doi: 10.1207/S15327906389-420

Masgoret, A-M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta–analysis of studies conducted by Gardner and associates. *Language Learning, 53,* 123–163. doi: 10.1111/1467-9922.00227

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

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide. Seventh edition*. Los Angeles, CA: Muthén & Muthén.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). Meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79,* 1129–1167. doi: 10.3102/0034654309337522

Möller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension of the internal/exernal frame of reference model to two verbal and numerical domains. *European Journal of Psychology of Education, 21*, 467–487. doi: 10.1007/BF03173515

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the „x" out of expectancy-value theory?: A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science*, *22*, 1058–1066. doi: 10.1177/0956797611415540

National Center for Educational Statistics (2015). *NAEP Foreign Language Assessmen*t. Retrieved from: http://nces.ed.gov/nationsreportcard/foreignlang/

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist, 41*, 181–206. doi:10.1207/s15326985ep4103_4

Pajares, F., Miller, M. D., & Johnson, M. J. (1999). Gender differences in writing self-beliefs of elementary school students. *Journal of Educational Psychology, 91*, 50–61. doi:10.1037/0022-0663.91.1.50

Pajares, F., & Valiante, G. (2001). Gender differences in writing motivation and achievement of middle school students: A function of gender orientation? *Contemporary Educational Psychology, 26*, 366–381. doi: doi:10.1006/ceps.2000.1069

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, *34*, 29–48. doi: 10.1080/01443410.2013.797339

Raykov, T. (2009). Evaluation of scale reliability for unidimensional measures using latent variable modeling. *Measurement and Evaluation in Counseling and Development, 42,* 223–232. doi: 10.1177/0748175609344096

Retelsdorf, J., & Köller, O. (2014). Reciprocal effects between reading comprehension and spelling. *Learning and Individual Differences, 30*, 77–83. doi: 10.1016/j.lindif.2013.11.007

Rhemtulla, M., & Little, T. D. (2012). Planned missing data designs for research in cognitive development. *Journal of Cognition and Development*, *13*, 425–438. doi: 10.1080/15248372.2012.717340

Rupp, A. A., Vock, M., Harsch, C., & Köller, O. (2008). *Developing standards-based assessment tasks for English as a first foreign language – Context, processes, and outcomes in Germany*. Münster: Waxmann.

Sawaki, Y., Stricker, L., & Oranje, A. (2008). Factor structure of the Toefl internet-based test (ibt): Exploration in a field trial sample. *ETS Research Report Series*, *2008*(1), 1–67. doi:10.1002/j.2333-8504.2008.tb02095.x

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177. doi: 10.1037/1082-989X.7.2.147

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46,* 407–441. doi: 10.3102/00346543046003407

Shell, D. F., Colvin, C., & Bruning, R. H. (1995). Self-efficacy, attribution, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement-level differences. *Journal of Educational Psychology*, *87*, 386–398. doi: 10.1037/0022-0663.87.3.386

Skaalvik, E. M., & Rankin, R. J. (1990). Math, verbal, and general academic self-concept: The internal/external frame of reference model and gender differences in self-concept structure. *Journal of Educational Psychology, 82*, 546–554. doi: 10.1037/0022-0663.82.3.546

Skaalvik, E. M., & Rankin, R. J. (1995). A test of the Internal/External Frame of Reference Model at different levels of math and verbal self-perception. *American Educational Research Journal, 32*, 161–184. doi: 10.3102/00028312032001161

Swann Jr., W. B., Chang-Schneider, C., & Larsen McClarty, K. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist, 62*, 84–94. doi:10.1037/0003-066X.62.2.84

Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology, 98,* 438–456. doi: 10.1037/0022-0663.98.2.438

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*, 111–131. doi: 10.1207/s15326985ep3902_3

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus.* Chichester, UK: John Wiley & Sons.

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

Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th-through 11th-grade Australian students. *Child Development, 75*, 1556–1574. doi:10.1111/j.1467-8624.2004.00757.x

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68–81. doi:10.1006/ceps.1999.1015

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89,* 451–469. doi: 10.1037/0022-0663.89.3.451

Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology, 89*, 420–432. doi: 10.1037/0022-0663.89.3.420

Wilgenbusch, T., & Merrell, K. W. (1999). Gender differences in self-concept among children and adolescents: A meta-analysis of multidimensional studies. *School Psychology Quarterly, 14*, 101–120. doi: 10.1037/h0089000

Williams, M., Burden, R., & Lanvers, U. (2002). 'French is the language of love and stuff': Student perceptions of issues related to motivation in learning a foreign language. *British Educational Research Journal, 28*, 503–528. doi: 10.1080/0141192022000005805

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER Conquest Version 2.0. Generalised item response modelling software*. Melbourne: Acer Press.

Xu, M. K., Marsh, H. W., Hau, K-T., Ho, I. T., Morin, A. S. J., & Abduljabbar, A. S. (2013). The internal/external frame of reference of academic self-concept: Extension to a foreign language and the role of language of instruction. *Journal of Educational Psychology, 105,* 489–503.doi: 10.1037/a0031333

Yeung, A. S., Chui, H-S., Lau, I. C-Y., McInerney, D. M., Russell-Bowie, D., & Suliman, R. (2000). Where is the hierarchy of academic self-concept? *Journal of Educational Psychology, 92*, 556–567. doi: 10.1037/0022-0663.92.3.556

Yeung, A. S., & Wong, E. K. P. (2004). Domain specificity of trilingual teachers' verbal self-concepts. *Journal of Educational Psychology, 96*, 360–368. doi: 10.1037/0022-0663.96.2.360

Table 1
Descriptive Statistics

			Boys				Girls			
	Rel.	α	*M*	*SD*	Skew	Kurtosis	*M*	*SD*	Skew	Kurtosis
Self-Concept										
G.SC.global	.80	.80	3.11	0.61	-0.48	-0.07	3.34	0.56	-0.63	-0.04
G.SC.S	.76	.69	3.36	0.53	-0.80	0.69	3.41	0.47	-0.64	0.13
G.SC.L	.74	.72	3.62	0.47	-1.54	3.27	3.74	0.38	-1.77	4.29
G.SC.W	.84	.84	2.96	0.69	-0.20	-0.57	3.21	0.66	-0.52	-0.31
G.SC.R	.83	.82	3.14	0.63	-0.55	-0.06	3.24	0.58	-0.57	-0.22
E.SC.global	.89	.89	2.99	0.72	-0.44	-0.35	3.10	0.68	-0.53	-0.16
E.SC.S	.83	.83	2.79	0.67	-0.23	-0.19	2.81	0.64	-0.13	-0.25
E.SC.L	.84	.84	2.91	0.66	-0.30	0.02	3.01	0.62	-0.31	-0.01
E.SC.W	.87	.87	2.65	0.70	-0.04	-0.21	2.81	0.71	-0.14	-0.38
E.SC.R	.88	.88	2.86	0.70	-0.30	-0.11	2.91	0.67	-0.18	-0.43
F.SC.global	.92	.93	2.62	0.86	-0.11	-0.79	2.92	0.81	-0.43	-0.55
F.SC.S	.89	.89	2.50	0.79	0.08	-0.57	2.67	0.76	-0.16	-0.50
F.SC.L	.91	.91	2.48	0.82	0.10	-0.67	2.74	0.76	-0.24	-0.41
F.SC.W	.92	.92	2.42	0.83	0.10	-0.65	2.75	0.81	-0.20	-0.56
F.SC.R	.92	.92	2.49	0.82	0.06	-0.58	2.72	0.78	-0.10	-0.50
Achievement										
G.ACH.L	.80		0.01	0.72	0.01	-0.27	0.04	0.88	-1.50	7.33
G.ACH.R	.74		-0.04	1.25	-0.16	0.90	0.07	1.21	0.05	0.79
E.ACH.L	.83		-0.05	1.69	-0.26	0.52	0.15	1.61	-0.40	0.90
E.ACH.R	.92		-0.14	1.47	-0.22	0.67	0.19	1.41	-0.59	1.79
F.ACH.L	.90		-0.14	1.26	-0.03	1.96	0.13	1.17	0.38	1.68
F.ACH.R	.84		-0.23	1.06	-0.09	0.38	0.13	0.93	-0.06	0.53
G.grade			3.11	.85	-0.16	-0.25	2.64	0.86	0.13	-0.26
E.grade			3.18	.91	0.01	-0.37	2.84	0.89	0.14	-0.33
F.grade			3.11	.95	-0.10	-0.22	2.63	0.93	0.20	-0.38

Note. SC = self-concept, G = German, E = English, F = French, S = speaking, L = listening, W = writing, R = reading, ACH = achievement, α = Cronbach's alpha reliability coefficient, Rel: Reliability [for self-concept: scale reliability (ρ; Raykov, 2009); for achievement: WLE reliability (Wu, Adams, Wilson, & Haldane, 2007)]. The self-concept means and standard deviations refer to manifest aggregations of the items.

Table 2

Goodness-of-fit Indices for First-order and Higher-order Factor Models

	Model description	χ^2	*df*	CFI	TLI	RMSEA	SRMR
1	Unidimensional self-concept re skills + global self-concept for German	4810.879	89	.775	.735	.112	.076
2	Unidimensional self-concept re skills + global self-concept for English	4102.505	89	.889	.869	.103	.048
3	Unidimensional self-concept re skills + global self-concept for French	2046.063	89	.900	.882	.113	.042
4	5-factor model for German	1421.118	80	.936	.916	.063	.044
5	5-factor model for English	965.406	80	.975	.968	.051	.021
6	5-factor model for French	395.685	80	.984	.979	.048	.016
7	Higher-order model for German + global self-concept	1725.941	85	.922	.904	.067	.057
8	Higher-order model for English + global self-concept	1119.941	85	.971	.965	.054	.025
9	Higher-order model for French + global self-concept	529.773	85	.977	.972	.055	.023
10	5-factor model for German + reading and listening tests	1527.825	100	.937	.915	.058	.041
11	5-factor model for English + reading and listening tests	1068.624	100	.975	.967	.048	.021
12	5-factor model for French + reading and listening tests	436.517	100	.985	.979	.044	.015
13	Higher-order model for German + global self-concept + school grade in German	1916.876	98	.920	.902	.066	.056
13a	Higher-order model for German + global self-concept + school grade in German; invariant correlations for higher order factor and global factor to school grade in German	2122.055	99	.911	.892	.069	.071
14	Higher-order model for English + global self-concept + school grade in English	1251.780	98	.969	.962	.053	.026
14a	Higher-order model for English + global self-concept + school grade in English; invariant correlations for higher-order and global factors to grade in English	1349.316	99	.967	.960	.054	.035
15	Higher-order model for French + global self-concept + school grade in French	617.526	98	.975	.970	.055	.026

[Table 1 continues.]

[Table 1 continued.]

	Model description	χ^2	*df*	CFI	TLI	RMSEA	SRMR
15a	Higher-order model for French + global self-concept + school grade in French; invariant relations	725.387	99	.970	.964	.060	.048
16	Higher-order model for German + global self-concept + combined achievement test scores	1919.607	112	.921	.903	.062	.054
16a	Higher-order model for German + global self-concept + combined achievement test scores; invariant relations for higher order and global factors	1922.162	113	.920	.904	.061	.054
17	Higher-order model for English + global self-concept + combined achievement test scores	1260.917	112	.971	.965	.049	.025
17a	Higher-order model for English + global self-concept + combined achievement test scores; invariant relations for higher order and global factors	1266.126	113	.971	.965	.049	.025
18	Higher-order model for French + global self-concept + combined achievement test scores	635.174	112	.976	.971	.052	.023
18a	Higher-order model for French + global self-concept + combined achievement test scores; invariant relations for higher order and global factors	643.070	113	.976	.971	.052	.024

Note. CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Squared Residual.

Table 3

Standardized Factor Correlations for the 5-factor Models (Models 4-6 in Table 2)

		Reading	Listening	Writing	Speaking
Listening					
	German	.635			
	English	.839			
	French	.882			
Writing					
	German	.642	.528		
	English	.729	.680		
	French	.801	.738		
Speaking					
	German	.621	.758	.605	
	English	.821	.865	.757	
	French	.889	.900	.824	
Global					
	German	.678	.550	.768	.640
	English	.860	.814	.787	.893
	French	.884	.840	.868	.919

Note. All correlations are statistically different from zero ($p < .001$).

Table 4

Standardized Correlations between Self-concept Facets and Achievement Test Scores for the 5-factor Models (Models 10, 11, 12 in Table 2)

	Reading self-concept	Listening self-concept	Speaking self-concept	Writing self-concept	Global self-concept
Reading test					
German	.449	.349	.319	.297	.305
English	.472	.471	.454	.427	.471
French	.520	.489	.488	.478	.512
Listening test					
German	.294	.292	.236	.194	.193
English	.443	.447	.426	.379	.409
French	.549	.574	.531	.474	.525

Note. All correlations are statistically different from zero ($p < .001$).

Table 5

Standardized Achievement Correlations for the Higher-order Factor Models (Models 13-18 in Table 2)

	Higher-order factor	Global factor
German grade	.457	.565
English grade	.480	.544
French grade	.566	.657
German combined test	.500	.342
English combined test	.568	.510
French combined test	.651	.602

Note. All correlations are statistically different from zero ($p < .001$).

Table 6
Goodness-of-fit Indices of Invariance Tests across Gender

	Model description	χ^2	*df*	CFI	TLI	RMSEA	SRMR
German							
I1	Configural invariance	1678.468	170	.927	.910	.065	.056
I2	Invariance of first-order factor loadings	1785.870	180	.923	.910	.065	.073
I3	Invariance of first-order and second-order factor loadings	1801.872	183	.922	.910	.065	.077
I4	Invariance of first-order and second-order factor loadings + testing invariance of first-order factors means	1986.859	193	.913	.906	.066	.084
I5	Invariance of first-order and second-order factor loadings + free estimation of second-order factor mean	2077.802	196	.909	.903	.067	.086
English							
I6	Configural invariance	1204.702	170	.971	.965	.054	.026
I7	Invariance of first-order factor loadings	1230.306	180	.971	.966	.052	.027
I8	Invariance of first-order and second-order factor loadings	1238.268	183	.971	.966	.052	.029
I9	Invariance of first-order and second-order factor loadings invariance of first-order and second-order factor loadings + testing invariance of first-order factors means factors	1386.746	193	.967	.964	.054	.031
I10	Invariance of first-order and second-order factor loadings + free estimation of second-order factor mean	1462.080	196	.965	.962	.055	.031
French							
I11	Configural invariance	615.083	170	.977	.972	.055	.024
I12	Invariance of first-order factor loadings	635.935	180	.976	.973	.054	.027
I13	Invariance of first-order and second-order factor loadings	640.336	183	.976	.973	.054	.027
I14	Invariance of first-order and second-order factor loadings invariance of first-order and second-order factor loadings + testing invariance of first-order factors means factors	670.037	193	.975	.973	.054	.028
I15	Invariance of first-order and second-order factor loadings + free estimation of second-order factor mean	705.016	196	.974	.972	.055	.030

Note. CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Squared Residual.

Supplementary Materials for

"Self-concepts in reading, writing, listening, and speaking: A multidimensional and hierarchical structure and its generalizability across native and foreign languages"

Table S1

Items including their Means (M) and Standard Deviations (SD, in parentheses) of the Various Self-concept Scales

	German	English	French
Global self-concept			
I am good at [German/English/French].	3.17 (0.686)	3.03 (0.748)	2.81 (0.885)
I learn things quickly in [German/English/French].	3.14 (0.719)	3.10 (0.790)	2.87 (0.893)
[German/English/French] is easy for me.	3.40 (0.705)	3.04 (0.795)	2.75 (0.934)
Reading self-concept			
Reading texts is easy for me.	3.49 (0.662)	3.14 (0.715)	2.83 (0.838)
I can quickly understand the content of texts.	3.23 (0.699)	2.95 (0.769)	2.70 (0.871)
I can quickly read and understand even difficult texts.	2.88 (0.754)	2.58 (0.794)	2.36 (0.879)
Listening self-concept			
When somebody talks in [German/English/French], I can easily understand.	3.85 (0.414)	3.10 (0.730)	2.74 (0.918)
It is easy for me to understand extensive spoken texts (e.g. audiobooks).	3.51 (0.632)	2.78 (0.765)	2.44 (0.828)
I can easily follow conversations.	3.70 (0.521)	3.02 (0.715)	2.75 (0.845)
Writing self-concept			
I am good at writing texts	3.05 (0.771)	2.72 (0.777)	2.62 (0.881)
I can express myself well in writing.	3.10 (0.771)	2.77 (0.775)	2.65 (0.882)
I do not have any difficulties writing a long text.	3.16 (0.815)	2.73 (0.839)	2.60 (0.943)
Speaking self-concept			
Talking in [German/English/French] is easy for me.	3.81 (0.459)	3.09 (0.769)	2.79 (0.898)
During discussions in class, I am always able to express myself comprehensibly.	3.33 (0.682)	2.80 (0.753)	2.66 (0.854)
I always find the right words even in conversations about difficult topics.	3.03 (0.713)	2.51 (0.738)	2.37 (0.824)

Note. Because the language the items refer to is not explicitly mentioned in all items, the questionnaire includes the following introduction before each of the section with self-concept items specifying the language a specific item relates to: " The following items relate to [German/English/French] language resp. the school subjects of [German/English/French]."