The dimensional structure of students’ self-concept and interest in science depends on course composition

Malte Jansen\textsuperscript{a,b,*}, Ulrich Schroeder\textsuperscript{c}, Oliver Lüdtke\textsuperscript{b,d}, Herbert W. Marsh\textsuperscript{e}

\textsuperscript{a} Institute for Educational Quality Improvement, Berlin, Germany
\textsuperscript{b} Centre for International Student Assessment, Germany
\textsuperscript{c} Department of Psychology, University of Kassel, Germany
\textsuperscript{d} Leibniz Institute for Science and Mathematics Education, Kiel, Germany
\textsuperscript{e} Australian Catholic University, Sydney, Australia

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Academic self-concept
Interest
Integrated science teaching
Dimensionality
Construct differentiation

\textbf{ABSTRACT}

Both academic self-concept and interest are considered domain-specific constructs. Previous research has not yet explored how the composition of the courses affects the domain-specificity of these constructs. Using data from a large-scale study in Germany, we compared ninth-grade students who were taught science as an integrated subject with students who were taught biology, chemistry, and physics separately with regard to the dimensional structure of their self-concepts and interests. Whereas the structure of the constructs was six-dimensional in both groups (self-concept and interest factors for biology, chemistry, and physics), the correlations between the domain-specific factors were higher in the integrated group. Furthermore, the pattern of gender differences differed across groups. Whereas male students generally showed higher self-concept and interest in physics and chemistry, a small advantage for male students in biology was only present in integrated science teaching group. We conclude that aspects of the learning environment such as course composition may affect the dimensional structure of motivational constructs.

1. Introduction

Academic self-concepts and academic interests are understood as domain-specific constructs; that is, students feel more confident and are more interested in some academic domains than in others. This theoretically assumed domain-specificity has been practically operationalized by measuring self-concept and interest separately for different school subjects (e.g., mathematics self-concept, English self-concept, etc.). However, school subjects vary by grade level, school type, and country. For example, science domains such as physics and biology can be taught either as part of an integrated science class or as separate school subjects. Researchers have yet to explore how such organizational characteristics affect the dimensional structure of domain-specific self-concepts and interests.

In this study, we make use of a natural experiment in Germany to close this research gap. Following the approach in other countries, integrated science teaching instead of the traditional approach of teaching biology, chemistry, and physics separately is now increasingly offered in Germany. Using a nationally representative dataset of ninth-grade students, we analyzed self-concept and interest measures in biology, chemistry, and physics. We compared students who were taught these domains as separate school subjects with students who were taught science as an integrated subject regarding (a) the dimensional structure of their academic self-concept and interest as well as (b) the pattern of gender differences in self-concepts and interests. We expected the students who were taught science as an integrated subject to show a less differentiated factor structure for both self-concept and interest.

2. Theoretical background and state of research

2.1. The structure of academic self-concept: development and differentiation

Shavelson, Hubner, and Stanton (1976) described academic self-concept as a multidimensional and domain-specific construct (e.g., academic self-concept in English, in mathematics, etc.). Since then, a plethora of empirical research has shown that such domain-specific facets can be distinguished (Brunner et al., 2010; Marsh, 1990). In the context of school education, the abovementioned domain-specificity has been operationalized by measuring self-concepts in different school
subjects. However, which science subjects are taught in a given learning environment varies substantially: Whereas in some countries such as the US the focus is on integrated science teaching, in other countries such as Germany physics, chemistry, and biology are usually taught separately. Accordingly, some educational large-scale studies such as the PISA studies have examined students' self-concepts in general science, whereas others have used more differentiated measures related to the science subdomains. Such research has shown that students differentiate between their academic self-concepts in biology, chemistry, and physics and that these facets are only moderately related (Hardy, 2014; Jansen, Schroeders, & Lüdtke, 2014).

Comparison processes based on achievement feedback are considered an important source of students' academic self-concept (Marsh et al., 2018; Wolff, Helm, Zimmermann, Nagy, & Möller, 2018). Students compare their current achievement in a domain with the achievement of their peers in the same domain (social comparisons), their own past achievement in the same domain (temporal comparisons), and their own achievement in other domains (dimensional comparisons). These comparison processes can explain individual differences in academic self-concept as well as the finding that students with similar achievement have diverging domain-specific self-concepts (Retelsdorf, Köller, & Möller, 2014; Wolff et al., 2018). However, comparison processes have mainly been studied with regard to their effect on the mean level of self-concepts rather than the dimensional structure of self-concept.

It has been suggested that the structure of self-concept should become more differentiated over the course of students' school careers offering as students receive more domain-specific achievement feedback (Shavelson et al., 1976; Stipek & Mac Iver, 1989). This idea is consistent with empirical evidence that different self-concepts tend to become more reliable and distinct with age (Marsh & Ayotte, 2003). Nevertheless, already primary school students in Grade 1 have been found to differentiate between a mathematical and verbal self-concept that become even more differentiated across primary school (Ehm, Lindberg, & Hasselhorn, 2014; Marsh & Ayotte, 2003; Schmidt et al., 2017). In contrast, Arens and Morin (2016) did not find differences in the correlational pattern between math and German across Grades 3 to 6, thus, hypothesizing that, for these two specific domains, the “differentiation process is already completed” (p. 22). In general, there is little evidence for further differentiation of academic self-concept after primary school, especially in the domain of science. The question which role structural characteristics such as course composition play in the development of the structure of domain-specific self-concept has not yet been addressed.

2.2. The structure of academic interests: development and differentiation

According to expectancy-value theory (Wigfield & Eccles, 2000), achievement motivation is affected not only by self-beliefs about one's competence (i.e., what students think they can do) but also by value beliefs (i.e., what students like to do). The most prototypical value belief is intrinsic value or individual interest. A student with high individual interest would be characterized by a strong cognitive commitment and emotional attachment to a specific domain (Krap, 2002). Thus, individual interest is assumed to be inherently domain-specific (Hidi & Renninger, 2006; Krap, 2002).

Compared with research on academic self-concepts, structural models of academic interest across academic domains have been less prominent. Results from previous research in an educational context confirmed the assumption of domain-specificity and, in the domain of science, have also suggested multidimensionality (Jansen, Lüdtke, & Schroeders, 2016): Similar to self-concepts, students held different levels of interest in biology, chemistry, and physics with the latter two more closely related. That the factor structure seems to be similar to the structure of academic self-concepts is not surprising given that the two constructs are substantially correlated and have been shown to co-evolve during the school career (Archambault, Eccles, & Vida, 2010; Denissen, Zarrett, & Eccles, 2007). Thus, they also show quite similar empirical patterns in their relations to other constructs even though the relation between achievement and self-concept is stronger than relation between achievement and interest (Trautwein et al., 2012).

Even though the constructs are empirically strongly intertwined and are assumed to co-predict achievement motivation following expectancy-value-theory (Wigfield & Eccles, 2000), their theoretical foundations differ: While the idea of achievement feedback and comparison processes as sources of interest has been less prominent (for an exception, see Schurz, Pfost, Nagengast, & Artek, 2014), more emphasis was put on the developmental part of individual interest from initial situational interest (e.g., the Four-Phase Model of interest development by Hidi & Renninger, 2006). One key assumption of this model is that the structure of interest becomes more elaborated over time and that interest becomes more differentiated across a student's school career. Hidi and Renninger also emphasized the importance of the learning environment for providing opportunities for students to engage with the content because individual interest can develop only through continuous engagement. The effect of course composition on interest development has not been studied so far.

2.3. The "gender gap" in science motivation

Gender differences in educational choices related to the STEM domains (Science, Technology, Engineering, and Mathematics) are still a matter of concern in educational systems worldwide (OECD, 2015). Compared with men, women are less likely to pursue higher education in STEM fields (Wang, Eccles, & Kenny, 2013) even though no substantial gender differences have been found in mathematics and science achievement (Hyde, Lindberg, Linn, Ellis, & Williams, 2008). A closer look, however, reveals differences across scientific domains (Cheryan, Ziegler, Montoya, & Jiang, 2017). Even given comparable school achievement, only 22% of first-year university physics students in Germany are female, compared with 43% female first-year students in chemistry and even 62% in biology (Schroeders, Penk, Jansen, & Pant, 2013). In the same vein, Wang and Degol (2013, p.20) concluded in a recent review examining factors that influence the ‘gender gap in STEM, that “little of this [previous] work has focused on the different occupational choices within STEM (e.g., physical sciences versus biological sciences)”.

Expectancy-value theory proposes differences in academic self-concepts and value beliefs as a central driver of differences in educational choices (Eccles, 2011; Wang & Degol, 2013). This assumption has been strengthened by studies showing that self-concepts and value beliefs affect both aspirations in relation to a career in a STEM field (Guo, Marsh, Parker, Morin, & Dicke, 2017; Nagengast et al., 2011) and actual educational choices (Parker et al., 2012). There is also evidence for gender differences in these facets of motivation. Advantages for male students in self-concepts and interests in mathematics have consistently been reported to emerge as early as the beginning of elementary school (Eccles, Wigfield, Harold, & Blumenfeld, 1993; Hyde, Fennema, Ryan, Frost, & Hopp, 1990; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002). In science, the pattern is more complex: There is a strong advantage for male students in physics, a smaller advantage in chemistry, and no gender difference in biology (Elster, 2007; Hardy, 2014; Jansen et al., 2014).

2.4. Integrated versus separated science teaching and its effect on the factorial structure of self-concept and interest

In Germany, science education in secondary school has a long tradition of separately teaching science subjects. Usually, biology is the first science subject taught from Grade 5 onwards, followed by physics and chemistry in Grades 7 and 8, respectively. To promote general scientific literacy, integrated science teaching is now increasingly
offered in Germany with the aim of emphasizing relations and commonalities between the different domains. In most cases, these integrated science subjects are offered in nonacademic track schools (that is, all school types other than Gymnasium; see Supplement A). But how does the teaching approach affect the development of the structure of self-concept and interest? Several potential mechanisms are conceivable:

First, the classes the students attend are labeled differently (“science” vs. “biology,” “chemistry,” “physics”) even though the general curriculum is quite similar (i.e., based on the national educational standards). For example, having 2 lessons of “biology,” 2 lessons of “chemistry,” and 2 lessons of “physics” each week rather than 6 lessons of “science” could have led students to develop separate motivational dispositions if these were indeed strongly aligned with the offered school subjects. The labels could, for example, trigger different stereotypes of students who do well in these subjects. Previous studies show that students develop such school-subject specific stereotypes and that their domain-specific self-concepts are influenced by self-prototype matching processes (Hannover & Kessels, 2004; Kessels, 2005).

Second, the school subject structure is also mirrored in grading. Students who are taught biology, chemistry, and physics separately also receive separate grades in each subject, whereas, in most cases, students who are taught integrated science receive only one grade. As grades are a very salient and important source of achievement feedback for students and this feedback, interpreted through comparison processes, is the base for students’ academic self-concepts (Jansen, Schroeders, Lüdtke, & Marsh, 2015; Möller & Marsh, 2013; Wolff et al., 2018), we would expect the structure of self-concept facets to mirror the structure of school subjects. More specifically, students compare and contrast their achievement across different domains if there is more than one facet of achievement feedback (dimensional comparisons; Möller & Marsh, 2013) as shown not only in a plethora of observational studies (Möller, Pohlmann, Köller, & Marsh, 2009), but also experimental studies based on manipulated achievement feedback (Möller & Köller, 2001; Strickhouser & Zell, 2015). Within the domain of science, such contrast effects have been shown between biology and physics (Jansen et al., 2015). The strength of contrast and assimilation effects varies depending on the perceived similarity between two subjects. Within the context of course composition, we think that the tailored achievement feedback that occurs when science domains are taught as separate subjects, will lead students to increasingly contrast their performance in those domains. In turn, this will result in a more differentiated factor structure with lower correlations between self-concept facets although their performance might be highly correlated (Jansen et al., 2014, 2015). Since dimensional comparison effects have also been found for interest (Schurtz et al., 2014) and students who feel competent also shown higher interest (Archambault et al., 2010), we would also expect a more differentiated interest structure to result from more differentiated grading.

Third, an integrated science subject is taught by a single teacher, whereas students are usually taught by different teachers who pursued different subject-specific courses in their university education. Teachers will differ in their style and quality of instruction, their grading, and their student support, et cetera. All this is likely to influence students’ self-concepts and interests (Kunter et al., 2013; Opdenakker, Maulana, & Brok, 2012). More specifically, different teachers might provide different opportunities for interest to develop, for example, by providing different levels of cognitive activation and support (Lazarides & Itel, 2012; Tsai, Kunter, Lüdtke, Trautwein, & Ryan, 2008), which might result in a more differentiated structure of interest.

Fourth, in addition to these more general mechanisms that are independent of the investigated domains, specific characteristics of integrated science teaching could lead to a less differentiated structure of self-concept and interest. For example, in integrated classes, there is a stronger focus on domain-general competencies such as scientific literacy, problem-solving skills, and the methods of scientific inquiry (Chowdhary, Liu, Yerrick, Smith, & Grant, 2014; Czerniak & Johnson, 2011). Moreover, integrated science teaching emphasizes the commonalities between the three science domains. For example, one popular method of integrated science teaching is to focus on one topic with a clear reference to everyday life and then discuss aspects of this topic from the perspectives of biology, chemistry, and physics. It should be noted again, however, that the national educational standards for science in secondary school are valid for all federal states independent of the composition of science courses and that generic competencies and including real-world problems is a pivotal goal of science teaching irrespective of the approach.

All four mechanisms point into the direction of stronger construct differentiation when science is taught in separate subjects. However, does this imply that there would be no differentiation at all for students attending integrated science classes? In general, people can possess self-concepts about areas for which there is no scholarly achievement feedback and that have no relation to academic domains (e.g., social self-concept or physical self-concept; Shavelson et al., 1976). In the academic domain, students also have distinguishable self-concepts about their abilities in the subdomains of school subjects. For example, they differentiate between reading, writing, listening, and speaking self-concepts in language domains, even though these skills are strongly related on an achievement level (Arens & Jansen, 2016). In summary, school subjects are not the least divisible unit of self-concepts but can be studied on a more fine-grained level. Contrary to self-concept research, interest research has more often moved beyond the level of school subjects to the level of more specific topics within domains such as science, and researchers have suggested that a distinction be made between interest in a school subject and interest in a particular topic—for example, students could be generally interested in topics in physics classes such as optics or astronomy but still show low interest in physics as a school subject as taught at a given time (Häussler & Hoffmann, 2000; Krapp & Prenzel, 2011). Thus, we would expect that students can differentiate between their self-concepts and interests in topics in the field of biology, chemistry and physics, even if they are taught within a single subject.

3. The present study

Using a nationally representative data set of German ninth-graders, we compared students who were taught science as an integrated subject from Grades 5 to 9 (integrated science teaching; abbreviated as “IST” in the following) with students who were taught biology, chemistry, and physics as separate school subjects (separated science teaching; abbreviated as “SST” in the following) with regard to the structure of their self-concepts and interests. Following the mechanisms outlined in the introduction, we hypothesized that a six-dimensional structure (self-concept and interest factors in biology, chemistry, and physics) would be found in both groups. However, we expected the self-concept and interest structures of SST students to be more differentiated compared to IST students. If the proposed mechanisms were indeed at work and resulted in stronger differentiation in the SST group, the factor correlations between the self-concept and interest facets in biology, chemistry, and physics should be higher in the IST group compared with the SST group.

As a second step of analysis, we studied gender differences. We expected to find a pattern of stereotypical gender differences in both groups as found in previous studies (Hardy, 2014; Jansen et al., 2014), that is, male advantages should be strongest in physics and still substantial in chemistry, whereas there should be no differences or small advantages for female students in biology. Given our hypothesis that
the self-concept and interest factors should be more differentiated when the domains are taught as separate school subjects, we would additionally expect that more differentiated factors would also go along with a stronger differentiation of gender differences. That is because, lower factor correlations would also allow for stronger differences in the relation between domain specific self-concept and interest factors and other variables such as gender. Based on this methodological argument, we would not expect a mean shift in the gender differences (i.e., the overall differences across the three domains becoming smaller or larger). In addition, it could be argued that characteristics of IST could lead to a mean shift and thus might help reduce gender differences. For example, teachers could more easily place a physical problem in a biological context and thereby raise female students’ interest in physics by tapping into their well-established topic-level interest in human biology and the environment (Beier & Ackerman, 2003; Elster, 2007). However, theoretically, such “contagion” effects could also work in the opposite direction and there are no previous studies on the effects of IST on gender differences that could inform a directed hypothesis.

4. Method

4.1. Study design

The data were derived from a large-scale assessment study focusing on students’ proficiency in math and science at the end of secondary education (IQB National Assessment Study 2012, see Lenski et al., 2016; Pant et al., 2015). The data were collected in 2012. The data set is one of the largest school assessment data sets available in Germany and consists among others of a students’ and school principals’ questionnaire. To determine the way science was taught, the principals of all participating schools were asked to specify how many weekly lessons of biology, chemistry, physics, and integrated science, respectively, were taught in each semester from Grade 5 to Grade 9. The principal questionnaire was mandatory for some of the federal states but not for all of them. A total of 890 school principals (67%) participated.

Participation was mandatory for the randomly sampled public schools, thus enabling a participation rate of 96.7% at the school level and 92% at the student level. The total sample consisted of 44,584 ninth-grade students from 1326 schools. The students were on average 15.02 years old (SD = 0.66); about half of them were female (49.5%). As is common in large-scale assessment, different test booklets and student questionnaires were randomly assigned to students (planned missing data design; Little, Jörgensen, Lang, & Moore, 2014; Rhemtulla & Little, 2012). The data were collected in spring 2012.

4.2. Measures

4.2.1. Academic self-concept

We assessed domain-specific self-concepts in biology, chemistry, and physics with four similarly worded items per domain: “I learn quickly in [biology/chemistry/physics],” “I have always believed that [biology/chemistry/physics] is one of my best subjects,” “I get good grades in [biology/chemistry/physics],” and “I am just not good at [biology/chemistry/physics]” (reverse scored). Students replied on a 4-point scale that had the response options strongly agree (coded as 4), agree (3), disagree (2), and strongly disagree (1). The items were taken from the international questionnaire used in PISA 2003 (originally for measuring self-concept in mathematics; Ramm, Adamsen, Neubrand, & Deutsches PISA-Konsortium, 2006). The version adapted for the three science domains was also used in a previous study (Jansen et al., 2015). The second and the third items may have been more difficult for the IST group to answer (because students usually received only one grade, and there were no separate subjects). However, the self-concept scales showed good reliability in both groups in our sample (coefficient $\omega > .85$, see Table 1), and they demonstrated measurement invariance across the two groups (see Table 1).

4.2.2. Interest

Interest was measured in biology, chemistry, and physics with items that had parallel wording: “I am interested in [biology/chemistry/physics],” “[biology/chemistry/physics] is important for me personally,” “Studying [biology/chemistry/physics] is fun,” and “Studying [biology/chemistry/physics] is one of my favorite activities. The response options, which used a 4-point scale, were strongly agree (coded as 4), agree (3), disagree (2), and strongly disagree (1). The items were taken from the German national questionnaire from the PISA studies (Ramm et al., 2006) and originally addressed interest in mathematics. The items refer to the cognitive and affective components of habitual, domain-specific interest (Krapp & Prenzel, 2011). The items were also used in a previous study on interest (Jansen et al., 2016). The reliability coefficients were very high in both groups ($\omega > 0.91$, see Table 1).

4.3. Sample of analysis and treatment of missing data

The total sample of the large scale assessment study consisted of 44,584 students. Using the data from the principal questionnaire and the participation rate of 67% mentioned above, there was information about the number of science lessons for only 32,512 students; all other students were excluded. Of these students, 13,307 received a version of the questionnaire that included self-concept items; 8040 questionnaires also included items about domain-specific interest. This was because the self-concept items were included in three of the eight versions of the questionnaires that were used, whereas interest was included in only two. Such planned missing data designs are often used in educational large-scale assessment as an economic and efficient way to collect data in a reasonable amount of time. The resulting dataset has data Missing Completely at Random (MCAR) which can easily be addressed with statistical techniques and does not induce bias (e.g., Enders, 2016; Little et al., 2014). For students who received a questionnaire that included the self-concept and interest items, respectively, the item-level rate of missing data was very low, ranging from 1.0% to 2.3%.

Because we aimed to analyze self-concept and interest jointly in a single model, we included all students who completed at least one self-concept item in the sample ($n = 12,967$). From this sample, students were further assigned to the SST or IST group. As an additional robustness check, we also replicated the results with other sample selection procedures (see Supplement G).

We estimated all models using Full Information Maximum Likelihood (FIML) estimation—a model-based approach for handling missing data. In the FIML procedure, missing data and parameter estimation are combined in a single step (Enders, 2010). FIML is considered superior to traditional methods of treating missing data because it allows for more precise parameter estimation with higher statistical power (Schafer & Graham, 2002). Furthermore, because it is unbiased under the MCAR and Missing At Random (MAR) assumptions, it is appropriate for planned missing data resulting from randomly distributed versions of a questionnaire.

4.3.1. Integrated teaching sample (IST)

Students were assigned to the integrated teaching sample if they had classes in integrated science from Grade 5 through Grade 9 and no classes in any of the separate science subjects during this period. As mentioned above, the form of integrated science teaching varies, and a mixed approach of integrated and separate science teaching is also common. However, not only would the many different forms have been hard to disentangle in our data set, but any mixed forms would also have made it harder to interpret possible group differences. Therefore, we decided to use the abovementioned strict criterion for the integrated teaching sample (a robustness check that included students with mixed forms of science teaching showed their result patterns were largely similar to the SST group, see Supplement E). This criterion was fulfilled by 337 students (of which 177 students completed both the self-concept and interest items, and 160 students completed only the self-concept
analyses using matched samples (see Supplement F). To ensure comparability between the IST and the SST group, we also replicated all the hierarchical structure of the data (students nested within classes), across the SST and IST groups. The con... items. As expected given the practice of integrated science teaching in Germany, all IST students were from nonacademic track schools.

4.3.2. Separate science teaching sample (SST)

Because only students from the nonacademic track could be identified for the IST sample, we also excluded all academic track students to ensure comparability. We assigned students to the separate teaching group if they had never received instruction in an integrated science subject from Grade 5 to Grade 9 and were thus taught biology, chemistry, and physics only as separate school subjects. This did not imply that SST students took every subject every semester but that they took each of the subjects at some time during secondary school. On the basis of this definition, we identified 4361 SST students.

4.4. Data analysis

4.4.1. Model estimation

We used a multigroup CFA framework to examine our central research question—whether the factor structure in the IST group would be less differentiated compared with the SST group. In both groups, the measurement model included six correlated first-order factors representing the academic self-concepts and interests in biology, chemistry, and physics. Furthermore, we tested whether there were differential mean gender differences in the IST versus the SST group. To be able to test these parameter differences (correlations, means), a series of measurement invariance tests were conducted. The procedure is described in detail in Supplement B. All models were estimated using the software Mplus 7.1. Correlated residuals between items with parallel wording across domains were estimated in all models (e.g., “I learn quickly in biology” and “I learn quickly in chemistry”). To account for the hierarchical structure of the data (students nested within classes), we corrected the standard errors. In order to check and ensure the comparability between the IST and the SST group, we also replicated all analyses using matched samples (see Supplement F).

5. Results

5.1. Invariance tests and factor correlations

After first examining descriptive statistics (see Supplement C), we tested our measurement model (six first-order factors representing self-concept and interest in the three domains) for measurement invariance across the IST and SST groups. The configurural invariance model showed a good fit to the data (CFI = .983, RMSEA = .030, SRMR = .028; see Table 1, Model 1). On the basis of this model, we consecutively fixed the factor loadings (weak invariance; Table 1, Model 2), the factor loadings and intercepts (strong invariance, Table 1, Model 3), and the factor loadings, intercepts, and residual variances (strict invariance, Table 1, Model 4) to be equal across groups. Even for the strict level of invariance, the decrease in model fit was very small (configural vs. strict invariance: Δ CFI = −0.001, Δ RMSEA = −0.001), indicating that strict measurement invariance could be established.

In the next step, we used the strict invariance model to further test for invariance in the factor variances and covariances using the χ²-difference test. The factor variances could be constrained to equality without a significant decrease in model fit (Δχ² = 6.92, df = 6, p = .33; see Table 1, Model 5). However, when the factor covariances were also constrained to equality, the χ²-difference test turned out to be significant (Δχ² = 136.41, df = 15, p < .01, see Table 1, Model 6).

Thus, as expected, there were significant and substantial differences in the pattern of correlations (see Table 2). All six constructs showed positive correlations in both groups. The correlations were higher within each domain (e.g., biology self-concept and interest in biology showed higher relations than biology self-concept and interest in chemistry) and within each construct (e.g., interest in biology and interest in chemistry showed higher relations than interest in biology and chemistry self-concept). For both self-concept and interest, the correlations between the domain-specific measures in biology, chemistry, and physics were much stronger in the IST group compared with the SST group as expected. That is, the correlations between the three self-concept facets ranged from .71 to .82 in the IST group, whereas the facets were only moderately correlated in the SST group (r = .30 to .45). This pattern was similar for interest with high relations between interest in biology, chemistry, and physics in the IST group (r = .66 to .87) and substantially lower relations in the SST group (r = .30 to .54).

5.2. Gender differences

After first establishing strong measurement invariance (for a description of the procedure, see Supplement B; for the model fit results, see Supplement D), we compared the factor means and the effect size (Cohen’s d) of the gender differences for the IST and SST groups. The results are shown in Fig. 1. In the SST group, we found a pattern of stereotypical gender differences that replicated previous results (e.g., Jansen et al., 2014): There were strong advantages for male students in physics (self-concept: d = 0.60, interest: d = 0.71), smaller but still

---

Table 1

<table>
<thead>
<tr>
<th>Model</th>
<th>S-B χ²</th>
<th>df</th>
<th>Δ S-B χ²</th>
<th>CFI</th>
<th>RMSEA</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Configural invariance model (six first-order factors; self-concept and interest in biology, chemistry, and physics)</td>
<td>1325.22</td>
<td>426</td>
<td>−</td>
<td>.983</td>
<td>.030</td>
<td>.028</td>
</tr>
<tr>
<td>2. Weak invariance model</td>
<td>1362.06</td>
<td>444</td>
<td>32.18</td>
<td>.982</td>
<td>.030</td>
<td>.030</td>
</tr>
<tr>
<td>3. Strong invariance model</td>
<td>1402.53</td>
<td>462</td>
<td>39.52*</td>
<td>.982</td>
<td>.029</td>
<td>.030</td>
</tr>
<tr>
<td>4. Strict invariance model</td>
<td>1438.70</td>
<td>486</td>
<td>41.98</td>
<td>.982</td>
<td>.029</td>
<td>.030</td>
</tr>
<tr>
<td>5. Strict invariance model with equal factor variances</td>
<td>1448.85</td>
<td>492</td>
<td>6.92</td>
<td>.982</td>
<td>.029</td>
<td>.032</td>
</tr>
<tr>
<td>6. Strict invariance model with equal factor variances and covariances</td>
<td>1639.61</td>
<td>507</td>
<td>136.41*</td>
<td>.978</td>
<td>.031</td>
<td>.053</td>
</tr>
</tbody>
</table>

Note. S-B χ² = Satorra-Bentler χ²; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual. n_SST = 4361; n_SST = 337. *p < .01.

---

Table 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.75*</td>
<td>.71*</td>
<td>.67*</td>
<td>.53*</td>
<td>.48*</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.75*</td>
<td>.82*</td>
<td>.52*</td>
<td>.71*</td>
<td>.53*</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.80*</td>
<td>.45*</td>
<td>.48*</td>
<td>.57*</td>
<td>.61*</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.69*</td>
<td>.25*</td>
<td>.19*</td>
<td>.66*</td>
<td>.68*</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.26*</td>
<td>.77*</td>
<td>.39*</td>
<td>.37*</td>
<td>.87*</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.19*</td>
<td>.35*</td>
<td>.76*</td>
<td>.30*</td>
<td>.54*</td>
<td></td>
</tr>
</tbody>
</table>

Note. The correlations are from a latent Confirmatory Factor Analysis model with strict measurement invariance and equal factor variances (Table 1, Model 5). n_SST = 4361; n_SST = 337. *p < .01.
substantial advantages for male students in chemistry (self-concept: $d = 0.31$, interest: $d = 0.32$), and only small differences in favor of female students in biology (self-concept: $d = -0.10$, interest: $d = -0.18$). In the IST group, there were still strong advantages for male students in physics (self-concept: $d = 0.56$, interest: $d = 0.52$), advantages in chemistry (self-concept: $d = 0.45$, interest: $d = 0.44$) and, most interestingly, also substantial advantages for male students in biology (self-concept: $d = 0.28$, interest: $d = 0.22$). A z-test revealed that only the effect size of the gender difference in biology (self-concept: $z = 2.853$, $p < .01$; interest: $z = 2.946$, $p < .01$) but not in chemistry and physics differed between the IST and SST groups. As an additional test of the robustness of this difference, we estimated a multiple indicator multiple cause (MIMIC) model instead of using a multigroup approach finding similar results (see Supplement H).

6. Discussion

We examined the effect of the course composition on the structure of students’ academic self-concept and interest in the sciences. Whereas both groups showed a domain-specific structure, the correlations between the three domain-specific self-concept and interest factors (biology, chemistry, physics) were considerably higher in the group that had received integrated science teaching instead of being taught biology, chemistry, and physics as separate school subjects. Furthermore, the pattern of gender differences in self-concept and interest differed between the groups. Whereas male students showed higher self-concept and interest than female students in physics and chemistry in both groups, their interest and self-concept in biology were marginally lower compared to female students in the SST group, but higher in the IST group. The results were very similar for academic self-concept and interest. Overall, the results suggest that course composition (i.e., the structure of school subjects that are taught) might significantly influence the differentiation of academic self-concepts and interests.

6.1. Implications for research on self-concept and interest

We believe these results contribute to the literature on the structure of self-concepts and interest in several ways. Using the natural experiment of IST versus SST as a case study is an innovative way to test effects of structural characteristics of the learning environment more broadly, and the course structure more specifically on the development of not only the means but the factor structures of motivational constructs. It has long been argued that domain specificity is a key characteristic of both self-concept and interest (Brunner et al., 2010; Hidi & Renninger, 2006; Krapp & Prenzel, 2011; Marsh, 1990; Shavelson et al., 1976). Our results show the central role of subject self-concept and interest seem to be linked to the specific ways in which curricula (i.e., differentiation of courses) and achievement feedback (i.e., differentiation of grading) are structured within schools. We outlined four—potentially interacting and cumulating—explanations for this effect: different course labels, differential grading, different teachers, and specific characteristics of integrated science teaching. Unfortunately, these cannot be tested in our study, but as will be outlined below, should be the focus of further studies. Still, we argue that first describing the phenomenon of course composition effects on the structure of motivational construct is an important first notion, especially since self-concept and interest are highly popular constructs in educational psychology and their domain-specificity has received so much attention in previous research.

Overall, we think the results point to the issue that even though the terms (school) “subject” and “domain” (e.g., “subject-specific self-
concept” or “domain-specific self-concept”) have largely been used interchangeably in the self-concept literature, they might mean different things in different school systems. Whereas the same domains (biology, chemistry, physics) were covered in both groups, the allocation of these domains to school subjects (one vs. three subjects) differed, and this difference seems to have had a strong effect even though the curriculum was quite similar between the groups. With regard to interest rather than self-concept, this idea has already been formulated. For example, in their comprehensive review of interest in science, Krapp and Prenzel (2011) described, “In order to distinguish between different kinds of science interests, it is obvious to refer to the structure of school subjects because these mainly provide the opportunities to get in touch with sciences systematically” (p. 33). However, they also pointed out that students can still be interested in different topics, contexts, and activities within and across school subjects. Based on this idea, more specific models have been developed to represent domain-specific interest (e.g., including many facets of interest in physics measured by as many as 88 items; Häussler & Hoffmann, 2000). Thus, in the interest literature, the level of school subjects has been described as only one of many levels that can be examined, whereas this point has not been stressed as much in the self-concept literature.

On a more general stance, our results show the importance of the learning environment for students’ responses to questionnaire items assessing self-concept and interest. One recommendation for future studies, and especially for cross-cultural studies, is that researchers should pay attention to the characteristics of the learning environment when studying domain-specific self-concepts and interests. For example, can students choose their courses and course levels? Or do all students receive the same amount of teaching? Is there implicit or explicit tracking (Chmielewski, Dumont, & Trautwein, 2013)? Can students choose not to attend courses in a given subject, and if so, how would the answers students give to self-concept or interest items be interpreted in that domain? How exactly are the courses labeled, and how do these labels compare with the items that are used? The answer to these more organizational questions which have been mostly neglected in previous research could have an influence on the structure of school-related psychological constructs. Science education is probably the most prominent example of differing course compositions across and even within countries. However, there are other situations in which course composition might play a crucial role. For example, in the US, math teaching has traditionally been strongly separated by topics such that entire courses of algebra, geometry, and pre-calculus are taught separately in separate school years. In most other countries, there is one integrated mathematics subject, and the topics change at a faster pace within each school year. The first case would be more likely to result in the students developing separate self-concepts or interests in different subdomains of math such as algebra and geometry compared with the second case.

6.2. Gender differences

Female students typically show lower self-concepts and interests in science than male students, particularly in chemistry and physics (Jansen et al., 2014). We replicated these results in our study. However, our focus was on comparing the pattern of gender differences in the IST and SST groups. We expected a less differentiated pattern of gender differences in the IST group compared with the SST group. On a descriptive level, whereas the effect sizes for the gender differences in the three domains were more similar in the IST group than in the SST group, this pattern was mostly driven by self-concept and interest in biology. In this domain, there were no gender differences in the SST group (a finding that is consistent with previous research on academic self-concepts; Hardy, 2014; Jansen et al., 2014). However, in the IST group, male students showed an unexpected advantage in biology. The gender differences in chemistry and physics were not statistically different across the course composition groups.

Regarding the discussion about gender gaps in STEM career choices, it seems difficult to decide whether a less differentiated factor structure of self-concept and interest and, correspondingly, a less differentiated structure of gender differences, would be desirable from a normative perspective. In our study, there was no significant reduction in female students’ disadvantages in physics or chemistry in the IST group, and there was even a newly established yet weaker disadvantage in biology. Thus, our study showed that IST does not generally reduce gender differences in self-concepts and interests in science.

We further hypothesized that IST might be a promising solution for closing the STEM gender gap and for strengthening female students’ self-concepts and interests if “contagion effects” from biology to the other two domains can be achieved. However, our results indicate that this mechanism, if at play at all, may have worked in the opposite direction. That is, the presence of chemistry and physics may have reduced the attractiveness of biology for girls in the IST group. Previous studies have shown that many students have an image of physics as a “masculine” subject and that the gender role of girls in many cases is at odds with showing high achievement in physics because students tend to ask whether their own self-concept lines up with the prototype of students who are good at physics (Hannover & Kessels, 2004; Kessels, 2005; Kessels & Hannover, 2008). It might be the case that students believe integrated science is a more masculine subject as well, and this belief may have reduced female students’ self-concepts and interest. However, future studies on the perception of integrated science would be required to test this hypothesis.

6.3. Limitations and directions for future research

Our study was based on cross-sectional data, and thus, we cannot make causal claims based on these data. However, we cannot think of possible confounding variables that are related to course composition (which is mostly defined by the federal state) and the structure of academic self-concepts and interests. Furthermore, we replicated all analyses with balanced samples (see Supplement F). Still, a longitudinal study of the effect of course composition would be desirable. This would allow researchers not only to make better causal claims but also to study the process of construct differentiation across different grades in more detail. For example, changes in the structure of self-concepts and interests when new science subjects are introduced (as mentioned above, in the SST group, biology is usually introduced first with chemistry and physics following in later grades) could be studied in more detail. Furthermore, it could be studied whether the composition of science courses affects not only self-concepts and interests but also future (gendered) educational choices.

In addition, even though our study used a large representative data set, the subsample of students matching our strict criteria for the IST group was much smaller than the SST group. As our additional analyses with other course composition groups showed, there appeared to be a qualitative difference between the strictly defined IST group and the groups that received IST in addition to SST (see Supplement E). Still, future studies should (over)sample this group to obtain a larger sample size.

Both in the Introduction and when discussing the theoretical implications, we proposed several mechanisms that might have contributed to the development of the differential structures of self-concept and interest. Unfortunately, owing to the cross-sectional approach and missing information on the specific practices across the science teaching groups, we could not test these mechanisms, which is a major limitation of our study. To fill this gap, future studies could, for example, apply a longitudinal design to determine whether the differentiation of constructs between the two groups slowly becomes stronger over time (i.e., test whether there are similar structures in the IST and SST groups in early secondary school that then become more distinct) or whether there are specific timepoints at which these differences are established (e.g., at the beginning of secondary school if the
composition of students in a course already differs at this point or when new school subjects are introduced). Furthermore, researchers could examine the developmental interplay between motivational constructs: For example, is self-concept affected first, and are the effects then transferred to interest? Or are both constructs simultaneously affected by course composition? To test an additional mechanism, students across the groups could be asked how similar they consider the different domains of science to be. Pointing out similarities is a goal of IST, and perceived similarity has been shown to affect the extent to which domain-specific self-concept is associated across different domains (Helm, Mueller-Kalthoff, Nagy, & Möller, 2016). In addition, this study could not explore how salient the domains “biology”, “chemistry” and “physics” were for students in the IST group given they don’t attend separated school subjects. Based on a comparison of topics covered in integrated science in early secondary school (Grasser, 2010), we know that there is a large variability from topics that can clearly be assigned to one domain only (e.g., “plants and animals to biology”), to others topics such as “the sun” or “water” that are clearly interdisciplinary. In future studies, IST students could be interviewed how they would characterize the domains of biology, chemistry and physics. Think-aloud protocols could also clarify how students respond to domain-specific self-concept and interest items and what information about the domains they use.

From the perspective of teaching, future studies could explore whether teachers who teach integrated science classes differ from single-subject teachers in their didactic approaches or classroom management. Regarding the effects of merely labeling a course or subject as well as the effects of grading, experimental vignette studies might offer a viable approach. From the perspective of science education, it would be interesting to compare the typical curricula (which vary by federal state and school track) in the IST and SST approaches with regard to the topics covered, the contexts, and the activities students engage in.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2018.11.001.

References

Among other changes, we extended the analyses to include interest in addition to self-concept, examine gender differences (using both MGCFA and MIMIC models) and introduced several robustness checks including, amongst others, propensity score matching (Appendix F) and examining two additional groups with different course structures (Appendix E).