Challenging the Universality of Achievement Goal Models: A Comparison of Two Culturally Distinct Countries

Anders Hofverberg & Mikael Winberg


To link to this article: https://doi.org/10.1080/00313831.2018.1544170

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 29 Nov 2018.

Submit your article to this journal

Article views: 12

View Crossmark data
Challenging the Universality of Achievement Goal Models: A Comparison of Two Culturally Distinct Countries

Anders Hofverberg and Mikael Winberg

Department of Science and Mathematics Education, Umeå University, Umeå, Sweden

ABSTRACT
Achievement goal theory is one of the most widespread motivation models within education research. Strong empirical support exists for the trichotomous model, comprising mastery-approach, performance-approach, and performance-avoidance goals. However, research also indicates problems with model transferability between contexts. In this study, based on questionnaire data from 4201 students, we use confirmatory factor analysis to compare the factor structures of students’ achievement goals in two culturally distinct countries. Factor structures for Grades 5–11 within the two countries were also compared. Results show that the separation between performance-approach and performance-avoidance goals differs between the two countries, and that this difference is consistent over the grades. Hence, results indicate that the model is not freely transferable between countries. The results are discussed in relation to differences in national culture and other proposed explanations such as age, perceived competence, and questionnaire characteristics.

For the last decades, achievement goal theory has been one of the most prominent theories of motivation in achievement situations, like education (Pintrich, 2003; Wirthwein, Sparfeldt, Pinquart, Wegerer, & Steinmayr, 2013). In general, there is a consensus that achievement goals comprise two main dimensions, performance and mastery goals. Performance goals are further divided into the valence-related avoidance and approach subdimensions, resulting in the trichotomous model including performance-approach (PAp), performance-avoidance (PAv), and mastery (considered an approach goal, MAp) goals. Despite strong support for this model (Elliot & Church, 1997), the correlation between PAp and PAv varies between studies, with several studies showing very high PAp–PAv correlations (Linnenbrink-Garcia et al., 2012). The variation in PAp–PAv correlations between studies leads to questions about the transferability of achievement goals models and, in cases of high PAp–PAv correlations, doubts about the practical significance of separating the two performance goals (Zusho et al., 2011).

Cultural influence is one of several suggested explanations for the variation in correlations between studies (Linnenbrink-Garcia et al., 2012). Still, existing research on achievement goals mostly assumes that constructs and models are freely transferable between contexts when comparing, for example, the level of goal adoption in different groups (e.g., Tang & Neber, 2008; Urdan, 2004) or the relations between achievement goals and outcomes in different cultures (e.g., Linnenbrink-Garcia, Tyson, & Patall, 2008, see also the meta-analysis by Hulleman, Schrager, Bodmann, & Harackiewicz, 2010). King and McInerney (2016) questioned this transferability assumption,
referring to the growth of sociocognitive, sociocultural, and situated theories of motivation, following an increased recognition of the influence of contextual factors on motivation. They argue that modern motivation research needs to acknowledge both universal and culturally specific features of motivation constructs. This argument is in line with Pintrich (2003), who stated that it is difficult to compare motivational beliefs, such as goals, between different cultural or ethnic groups without “considering the contextual and cultural meanings and functions of the constructs” (p. 681). In this study, we will investigate the factor structure of students’ achievement goals in Sweden and Germany, two culturally distinct countries, thus testing the transferability of achievement goal models between these two contexts. We will also compare achievement goal models over Grades 5–11 within each country, testing the stability of models over grades and thereby how representative the models are for students in each country.

Background

Achievement Goals

In this paper, we adopt Hulleman et al.’s (2010) definition of an achievement goal as “a future-focused cognitive representation that guides behavior to a competence-related end state that the individual is committed to either approach or avoid” (p. 423). Accordingly, the adoption of different achievement goals guides students to different learning behaviors. Students pursuing performance goals are characterized by a normative evaluation of competence, associated with a desire to demonstrate and validate competence in relation to others. In contrast, mastery goals are associated with an absolute, intrapersonal evaluation of competence with aims such as learning as much as possible, mastering tasks, improving, and gaining understanding (Ames, 1992; Elliot & McGregor, 2001).

The mastery and performance dichotomy has been developed into more elaborate models, initially through the addition of an approach and avoidance distinction for performance goals, resulting in a separation between performance-approach (PAp) and performance-avoidance (PAv) goals (see e.g., Elliot, 1999; Elliot & Church, 1997). The difference between an approach goal and an avoidance goal lies in the valence of the goals: seeing the chance of success and approaching it, or seeing the risk of failure and avoiding it (Elliot & Harackiewicz, 1996). Mastery goals have also been divided into mastery-approach (MAp) and mastery-avoidance (MAv) goals, forming a 2 × 2 goal framework (suggested in Elliot, 1999; and presented more thoroughly in Elliot & McGregor, 2001; and Pintrich, 2000a). Besides adding the MAv goal, the 2 × 2 framework excluded the standpoint subcomponent of achievement goals, that is, whether students aim at developing (mastery goals) or demonstrating (performance goals) competence. Consequently the 2 × 2 framework focus exclusively on the standard, or reference, for competence evaluation, that is, whether competence is evaluated by a task or self-based (mastery) or other-based (performance) standard (Elliot & Hulleman, 2017).

Although there is empirical support for the 2 × 2 framework (Elliot & McGregor, 2001; Elliot & Murayama, 2008), there are also doubts about the utility of a MAv goal (Bong, 2009), and concerns for its overlap with other goals, primarily PAv (Maehr & Zusho, 2009). Furthermore, the three goals PAp, PAv, and MAp “have produced the most solid empirical base” (S. Lau & Nie, 2008, p. 15), and the trichotomous model is consequently the model applied in the present study.

MAp goals have been shown to correlate positively with interest (Church, Elliot, & Gable, 2001), use of productive learning strategies (Wolters, 2004), positive emotions (Winberg, Hellgren, & Palm, 2014), and high effort (Wolters, 2004). PAv goals show overall unfavorable patterns of outcomes (Church et al., 2001; Elliot & Church, 1997; Elliot & McGregor, 2001; Wolters, 2004), whereas the results for PAp goals are more mixed. PAp goals have mostly showed positive correlations with academic achievement and interest (see e.g., Church et al., 2001; Elliot & McGregor, 2001; Hulleman et al., 2010; Wolters, 2004), but also negative patterns in relation to students’ effort, behavior, social relations in the classroom, and test anxiety (Kaplan, Gheen, & Midgley, 2002; S. Lau & Nie, 2008 Muis, Ranellucci, Franco, & Crippen, 2013; Polychroni, Hatzichristou, & Sideridis, 2012.)
Research using the trichotomous model has used both the standard and standpoint subcomponents, either separately or together (Elliot & Hulleman, 2017). Although this surely has contributed to the mixed results of PAp, Senko and Tropiano (2016) argue that neither approach is clearly superior on theoretical grounds and that studying them together should be a more fruitful approach, capitalizing on their respective strengths. Moreover, in the context of cross-cultural studies, the standpoint subcomponent of performance goals (to demonstrate competence) may be important to take into account potential differences in how self-promoting behaviors are viewed in different cultures, which will be discussed later in this paper. For these reasons, we have used the standard and standpoint subcomponents together for this study.

**Developmental Aspects of Achievement Goals**

In their review, Stipek and Iver (1989) found evidence that children’s conceptualizations of ability become more differentiated with age, and that the criteria children use for evaluating their competence changes over the elementary school years; from using social reinforcement, the amount of effort invested in the task/learning, and mastery, to a more objective and normative comparison. From a theoretical perspective, these changes in the way children assess their competence could affect what goals they can differentiate between. However, empirical studies show ambiguous results regarding the relationship between age and the structure of students achievement goals. Bong (2009) found that younger children (i.e., Grades 1–4) seemed to have more difficulty to discriminate between the goals in the $2 \times 2$ model than older students (Grades 5–9), with students from grades 1–2 showing particularly high correlations between PAp and PAv ($r = .58$) and MAv and PAv ($r = .69$). However, in their review, Linnenbrink-Garcia et al. (2012) found considerable variation in reported correlations between PAp and PAv goals; a “handful” (p. 286) of studies showing small negative or positive correlations ($- .30 < r < .10$), a majority of studies showing moderate correlations ($.30 < r < .50$) and “quite a few studies with correlations above .50” (p. 286). Citing the results from the meta-study by Hulleman et al. (2010), Linnenbrink-Garcia et al. (2012) concluded that there is little evidence that the developmental level can explain the variation in PAp–PAv correlations found in the literature. Despite this conclusion, we argue that the considerable diversity of results warrants further studies on this topic.

**Achievement Goals and Culture**

In this paper, we will discuss culture as a national-level phenomenon. Although a nation can comprise several different cultures, nations are considered to be “… the source of considerable amount of common mental programming of their citizens” (Hofstede, Hofstede, & Minkov, 2010, p. 21) and is therefore a suitable unit of analysis. Culture can be defined as shared “patterns of thinking, feeling, and acting” that distinguish groups of individuals from other groups (Hofstede et al., 2010, p. 5), and has a key role in understanding students’ motivation and learning (King & McInerney, 2016).

Previous studies on measurement invariance of achievement goal constructs across cultures show inconclusive results. McInerney and colleagues (McInerney & Ali, 2006; McInerney, Roche, McInerney, & Marsh, 1997; McInerney, Yeung, & McInerney, 2001) report invariance across cultural groups drawn from Australia, US, Canada, Hong Kong, and Africa, though they apply other achievement goal models than the most common two-goal, trichotomous, or $2 \times 2$ models. Despite the reported invariance, not all items are invariant across groups (McInerney et al., 2001). Furthermore, the confirmatory factor analysis (CFA) models presented in McInerney et al. (1997) show questionable fit in the separate cultural groups and the invariance claims in McInerney and Ali (2006) are difficult to evaluate because deterioration of fit between levels of invariance is not presented.

Using a two-goal mastery and performance model, with self-efficacy added as a third factor, Meissel and Rubie-Davies’s (2016) study of achievement goals in various cultural groups in New Zealand support strong invariance between the groups. Additionally, Murayama, Elliot, and Yamagata (2011)
showed measurement invariance between Japanese and US students for a performance goal model separating PAp and PAv, whereas Murayama, Zhou, and Nesbit (2009) claim to show measurement invariance between Japan and Canada for a 2 × 2 goal model. Although Murayama et al. (2009) claim invariance between Japan and the US, we consider some of their results to indicate otherwise (e.g., they report a ΔCFI of −0.02 for the metric model compared to the configural, exceeding the maximum of −0.01 recommended by Cheung & Rensvold, 2002). In conclusion, previous studies mainly support cross-cultural measurement invariance. However, large intergroup variability in PAp–PAv correlations have been observed, and high correlations between PAp and PAv are common, leading to uncertainty regarding the structure and measurement of the achievement goal model (Linnenbrink-Garcia et al., 2012). At a special conference session, dedicated to the issue of high PAp–PAv correlations, Zusho et al. (2011) even questioned the practical significance of models separating PAp and PAv. Hence, ambiguities regarding the applicability of the achievement goal model in different groups exist, warranting further investigation.

Culture has been suggested as a plausible explanation to the variation in PAp–PAv correlations, though clear evidence is missing (Linnenbrink-Garcia et al., 2012). Still, Zusho and Njoku (2007) showed that correlations between PAp and PAv were higher for Asian Americans than for Anglo Americans, and Bong, Woo, and Shin (2013) refer their problems in separating PAp and PAv to the Korean culture and the “threatening” school environment that it entails.

**Cultural Dimensions**

Many previous achievement goal studies lack a clear theoretical framework to describe cultures or an empirical measure of culture. Often, groups of students have been positioned along an arbitrary individualistic-collectivistic continuum, reflecting the degree of interdependence between the members of a group. Linnenbrink-Garcia et al. (2012) showed that the results on the relationship between culture and PAp–PAv correlations are ambiguous, arguing that part of this ambiguity might stem from variation in the definition and operationalization of culture between studies. There are many perspectives on how to define culture. However, within contemporary social research, culture is often framed as multidimensional; “entailing implicit and explicit patterns of meanings, values, and behaviors that identify the members of a cultural group” (Zusho & Clayton, 2011). In contrast, in motivation research, culture is often described from only one perspective, such as the individualistic-collectivistic dimension (Zusho & Njoku, 2007), and/or implicitly in terms of “Western” or “Eastern/Asian” culture (Murayama et al., 2009), nationality (Murayama et al., 2011), or ethnicity (Meisels & Rubie-Davies, 2016). To better understand the role of culture in the development of students’ achievement goals, there is a need for further studies on the topic (Linnenbrink-Garcia et al., 2012), and “researchers may [also] want to measure the critical cultural ingredients purported to be responsible for cross-cultural differences” … “variables that are never in fact actually tested” (p. 4; King & McInerney, 2016). Moreover, King and McInerney (2016) called for research including other dimensions of cultural variability than the individualistic-collectivistic. We argue that such research should not only include other measures of culture, but also more comprehensive (i.e., multidimensional) measures of culture to facilitate the identification of critical cultural ingredients for motivation. Although such identification is beyond the scope of this paper, we suggest that Hofstede’s Index of National Cultures (Hofstede et al., 2010), comprising six cultural dimensions, may offer a useful framework for describing culture. The six dimensions of the Index of National Cultures (as presented in Hofstede et al., 2010), and the associated characteristics relevant for education, are

- **Power distance** (large distance means teachers are considered as authorities above critique, teachers transfer their wisdom to students, and the quality of learning depends on the teacher).
- **Individualism versus collectivism** (individual opinions vs. preferences of the group, learning how to learn vs. learning how to do, learning is a lifelong process vs. a onetime “rite of passage” [p. 119]).
Masculinity versus femininity (competition vs. self-fulfillment, achievement in school is high stake vs. low stake. The labels masculinity and femininity are unfortunate as they draw attention to gender differences rather than the performance-related features of competitiveness and normative comparison that the dimension represents. Henceforth, we will refer to this dimension as competitiveness, with high masculinity score corresponding to high competitiveness).

Uncertainty avoidance (strong avoidance means low tolerance of ambiguity and open-ended learning situations, avoidance of intellectual disagreement with authorities, beliefs in right or wrong answers, and attributions of results to external and uncontrollable factors, like luck, rather than own ability).

Long-term versus short-term orientation (sustained effort vs. quick learning, adaptiveness vs. stability, dualism vs. relativism, concrete vs. abstract).

Indulgence versus restraint (optimism vs. pessimism, the perception of personal life control vs. perception of helplessness, freedom of speech is important vs. [relatively] unimportant, less moral discipline vs. moral discipline).

Of these dimensions, the competitiveness dimension should be highly relevant to achievement goal studies as it closely relates to performance goal structures. Performance goal structures are environmental factors that emphasize competition and the importance of success, and they relate to students’ adoption of performance goals (Ames, 1992). Similarly, a society with high competitiveness score is driven by competition and success defined in comparison with others (Hofstede et al., 2010), whereas in a society with low competitiveness, people tend to strive for self-fulfillment rather than competitive performance. In competitive school cultures, the best students are considered as role models and competition between students is promoted. Consequently, Hofstede et al. argue that students in highly competitive environments tend to be open with their competitive ambitions and try to make their competence visible. Moreover, attempts to excel are not only socially accepted, but encouraged, and successful students are admired and considered as role models for other students. However, in this context, failure often has negative implications for, inter alia, students’ social status in the classroom. In contrast, in low competitiveness school cultures, the average student is the norm and even weak students are praised (e.g., for their efforts to improve). There are also less dramatic consequences of failure. Moreover, in low competitive cultures, self-promoting behaviors and attempts to excel (cf., the performance goal standpoint subcomponent), especially on behalf of others, are generally not socially accepted.

The two countries in this study, Sweden and Germany, show a large difference in the competitiveness dimension (see Figure 1). Sweden is the least competitive country of all 76 regions listed and Germany one of the most competitive, close to USA (Hofstede et al., 2010) where much of previous achievement goal research has been conducted. Hence, it is plausible that the achievement goal model of the German students would be similar to those found in many American studies, that is, discriminating between MAP, PAP, and PAV goals. In contrast, the less competitive environment in Swedish schools, and less severe consequences of failure, could mean that students have less reason to discriminate between approach and avoidance goals. If there are few negative consequences of failure and few rewards for success, the difference between avoiding failure and approaching success may be less noticeable for the students than the standard and standpoint used to evaluate success. Thereby, students adopt a general performance goal without the approach–avoidance distinction. Consequently, achievement goal models may differ between Sweden and Germany.

Other noticeable differences between Sweden and Germany exist for the cultural dimensions of long-term orientation (LTO), uncertainty avoidance (UA) and indulgence. Since the differences between Sweden and Germany are smaller in these dimensions than in the competitiveness dimension, and since we find the theoretical links between those dimensions and students achievement goals less straightforward, they are only briefly discussed here. According to Hofstede et al. (2010), high LTO in learning situations is associated with a focus on the concrete (e.g., how to solve problems), self-improvement, and attribution of results to own effort (rather than external
and uncontrollable factors such as luck or the teacher). In contrast, low LTO implies a focus on the abstract, self-enhancement, and a tendency to dismiss information about oneself that implies low ability. In cultures with high UA, the trust in authorities, such as teachers, is high and there is a strong preference for structure in terms of rules, deadlines, and clear tasks with single unambiguous answers. We posit that, theoretically, high LTO could be associated with mastery goals, at least indirectly (focusing on self-improvement and making attributions to effort) whereas high UA could be associated with performance goals (preference for unambiguous answers, teachers as unquestionable bearers of knowledge). However, these theoretical connections are not as straightforward as for competitiveness.

Aim and Research Questions

This paper aims to investigate the universality of achievement goals by comparing the factor structures of students’ achievement goals in Sweden and Germany and between grades within the two countries respectively.

- Which achievement goal factor structure is most appropriate for describing chemistry achievement goals in Sweden and Germany, respectively?
- How consistent are the respective factor structures over Grades 5–11?

Methods

In this study, we collected cross-sectional questionnaire data from German and Swedish students in Grades 5–11. We then compared several possible achievement goal models through confirmatory factor analysis (CFA; Tabachnick & Fidell, 2013), trying to find which factor structure that was most appropriate for Swedish and German students, respectively. Next, we explored which factor structure that was most appropriate for each grade within the two countries. Finally, we investigated measurement invariance (MI) between groups, thereby testing if students in different groups conceptualized achievement goals in the same way.

Participants

In total, 2109 students (53% female) in Grades 5–11 from two different Swedish municipalities participated in the study. All Swedish students attended public schools, governed by the municipality. The German sample was collected from a single federal state and consisted of 3009 students (54%
female) in Grades 5–11. After data cleaning, 1470 Swedish students (55% female) and 2731 German students (55% female) remained for further analyses. For more details regarding data cleaning, see “Cleaning procedure” below. A majority (92%) of the German students attended Gymnasium, the highest track of German education, preparing students for higher studies. In the federal state where we conducted this study, the choice of academic track is ultimately the parents, but the primary school teachers provide a recommendation. In this state, of the students that started grade 5 in a public secondary school year 2015/2016, 42.8% started at Gymnasium, while 56.9% started the lower academic track, Gemeinschaftsschule (Landesregierung, Ministerium für Schule und Berufsbildung, 2017). The Swedish schools did not differentiate between students based on school achievement. However, no students from specific programs for students who had recently immigrated to Sweden or from programs for students with learning disabilities participated in the study. Grades 10 and 11 (upper secondary level) are not compulsory in Sweden, but most students (98% year 2015; Skolverket, 2016) transition directly from lower to upper secondary school, that is, from Grade 9 to Grade 10.

**Measures**

Due to the previously mentioned critique of the MAv goal, we used the MAp, PAp, and PAv subscales for the present study. To capture the standard and standpoint subcomponent of the trichotomous model, we combined items from the Revised Achievement Goal Questionnaire (AGQ-R, Elliot & Murayama, 2008) and the Patterns of Adaptive Learning Scales (PALS; Midgley et al., 2000).

During the development of the questionnaire, all scales were validated in four rounds (n = approximately 150 per round); including principal component analysis (PCA) of response patterns, interviews with students about items that did not load as expected, and revisions of problematic items between rounds. The validation studies showed that students did not differentiate between the AGQ-R items within the respective subscales and that the similarities between those items had a negative impact on students’ motivation to complete the questionnaire. Each subscale in the AGQ-R contains three items with similar meaning and phrasing. To reduce the repetitiveness, we changed the first part of the sentence of one item in each of the original AGQ-R subscales to “It is important to me to…” It could be argued that “importance” is an affective aspect, generally not considered a part of the achievement goal model. However, according to Hulleman et al. (2010) this wording “… assess or allow inference of goal-directed reasons or standards for achievement-relevant behavior …” (p. 431) and is in line with the definition of an achievement goal used in this paper.

Moreover, to increase the specificity of the context, the subject area (chemistry) was spelled out in most of the items and, in some items, performance was defined as achievement on tests, for example, “In chemistry it is important for me to perform better on tests than the other students” (Ag 1). For all items, students responded on a 5-point Likert scale. See Table 1 for a complete list of items and their origin.

The items were translated from English to Swedish and German, to be distributed to students in their countries’ native language. An independent researcher fluent in both languages validated the translations.

**Data Collection**

This study is part of a larger study that includes, besides achievement goals, measures of motivation type,\(^1\) attributions,\(^2\) epistemic beliefs,\(^3\) classroom perceptions,\(^4\) interest,\(^5\) self-concept,\(^6\) and

---

\(^1\)Adapted from SRQ-A (Ryan & Connell, 1989).

\(^2\)Based on Weiner (1985, 2000).

\(^3\)Adapted from Conley, Pintrich, Vekiri, and Harrison (2004).

\(^4\)Adapted from Autonomy support (Williams & Deci, 1996), perceptions of teacher goals (Midgley et al., 2000), What is Happening in the Classroom? (WIHIC; Aldridge & Fraser, 2000).

\(^5\)Adapted from Diers, Höfler, and Parchmann (2014)

\(^6\)Adapted from Marsh et al. (2008) and Spinath, Stiensmeier-Pelster, Schöne, and Dickhäuser (2002).
conceptual understanding of chemistry. Therefore, the instrument measuring achievement goals was distributed together with other instruments in a combined questionnaire. We collected data from Swedish students digitally via WebSurvey® and from German students through a paper-and-pen questionnaire. Using different data collection modes in Germany and Sweden introduced a possible source of differences between the countries, but Hox, De Leeuw, and Zijlmans (2015) concluded that most previous studies show strong invariance between different modes, especially between web-based and pen-and-paper questionnaires. Still, online questionnaires tend to produce more careless responses (defined as responses without regard to item content) than pen-and-paper questionnaires, which could threaten data quality (Meade & Craig, 2012). We address this problem in the next section, where we describe the procedure for cleaning data. To reduce the effects of fatigue for particular items, we distributed three versions of the questionnaire, with different item order, in approximately equal proportions to each class in both countries.

### Cleaning Procedure

Our initial data screening showed that several student responses contained long strings of identical answers. This, in turn, led to positive correlations between all constructs, even those hypothesized to be negatively correlated. Cleaning data from careless responses was thus deemed necessary and for this purpose, we decided on a procedure for calculating cutoff values for students’ maximum longest string of identical answers (Meade & Craig, 2012). We defined cutoff values as maximum longest string values exceeding 1.5 times the interquartile range above the upper quartile (i.e., outliers). Cutoff values were validated by examining scree plots of maximum longest string (see Johnson, 2005). Responses with longest string values exceeding the cutoff value were filtered out in all subsequent analyses.

### Table 1. Subscale items and their respective origin.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Item</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mastery-approach</td>
<td>Ag 5 It is important for me to understand chemistry as well as possible.</td>
<td>Reformulation of AGQ-R/AGQ: I am striving to understand the content of this course as thoroughly as possible.</td>
</tr>
<tr>
<td></td>
<td>Ag 6 I strive to develop a broad and deep knowledge in chemistry.</td>
<td>Reformulation of AGQ-R: I strive to completely master the material presented in this class.</td>
</tr>
<tr>
<td></td>
<td>Ag 7 My goal is to learn as much as possible in chemistry.</td>
<td>AGQ-R</td>
</tr>
<tr>
<td></td>
<td>Ag 8 In chemistry, I want to learn things, even if they are not assessed on tests or affect my grades.</td>
<td>New item</td>
</tr>
<tr>
<td>Performance-approach</td>
<td>Ag 1 In chemistry it is important for me to perform better on tests than the other students.</td>
<td>Reformulation of AGQ-R: My goal is to perform better than the other students.</td>
</tr>
<tr>
<td></td>
<td>Ag 2 In chemistry my goal is to perform better than other students.</td>
<td>Reformulation of AGQ-R: My goal is to perform better than the other students.</td>
</tr>
<tr>
<td></td>
<td>Ag 3 One of my goals is to show others that I am good at chemistry.</td>
<td>PALS: One of my goals is to show others that I’m good at my class work.</td>
</tr>
<tr>
<td></td>
<td>Ag 4 It is important to me that I look smart compared to others in my class.</td>
<td>PALS: It’s important to me that I look smart compared to others in my class.</td>
</tr>
<tr>
<td>Performance-avoidance</td>
<td>Ag 9 My goal is to avoid doing worse in chemistry than other students.</td>
<td>AGQ-R: My aim is to avoid doing worse than other students.</td>
</tr>
<tr>
<td></td>
<td>Ag 10 In chemistry, it is important for me to not perform worse than other students on tests.</td>
<td>AGQ-R: I’m striving to avoid performing worse than others.</td>
</tr>
<tr>
<td></td>
<td>Ag 11 In chemistry, it is important to me that I don’t look stupid.</td>
<td>PALS: It’s important to me that I don’t look stupid in class.</td>
</tr>
<tr>
<td></td>
<td>Ag 12 One of my goals in class is to avoid to show that I have trouble understanding</td>
<td>PALS: One of my goals in class is to avoid looking like I have trouble doing the work.</td>
</tr>
</tbody>
</table>

*Based on Hadenfeldt, Bernholt, Liu, Neumann, and Parchmann (2013).*
Statistical Analyses

First, we used IBM SPSS Statistics version 23 to test data factorability and check linearity and correlations between items, thus making sure data were fit for CFA. Second, we performed CFA using Mplus version 7.11 (Muthén & Muthén, 1998–2015). For all CFA models, each item was allowed to load freely on one latent factor, but had zero loadings on all other factors; latent factor variances were set to 1.0; and latent factors were allowed to correlate.

We compared five CFA models, representing all possible combinations of MAp, PAp, and PAv goals:

Three-factor model:
(A) MAp–PAp–PAv

Two-factor models:
(B) MAp–PAp/PAv (Mastery–Performance)
(C) Map/PAp–PAv (Approach–Avoidance)
(D) MAp/PAv–PAp

One-factor model:
(E) MAp/PAp/PAv

We fitted these five models to the data for Sweden and Germany, respectively, and also separately for each grade within the two countries.

Data were ordinal (5-point Likert scale), but were treated as continuous in CFA, as is common practice (Wang & Wang, 2012). However, ordinal data by definition violate assumptions of multivariate normality. Therefore, we used Mplus’ robust maximum likelihood (MLR) estimator that is robust against violations of normality and able to handle missing data (Muthén & Muthén, 1998–2015; Wang & Wang, 2012).

To evaluate the fit of the CFA models, we focused on four goodness-of-fit indices (GFIs): root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), comparative fit index (CFI), and Tucker–Lewis index (TLI). Acceptable value for both RMSEA and SRMR is <.08 (Browne & Cudeck, 1992; Hu & Bentler, 1998), even if RMSEA < .06 is more desirable (Hu & Bentler, 1999). For CFI and TLI, cut-offs are generally >.90, as models with values <.90 often can be significantly improved (see Bentler & Bonett, 1980), but >.95 would be more desirable (Hu & Bentler, 1999). Furthermore, we considered model parsimony in evaluating what the best model was, as a less complex model is preferable when two models show equally good fit (Kline, 2016).

If the fit of the model did not reach the cutoff values, carefully judged post hoc modifications based on modification indices were applied (see e.g., Byrne, 1989; Sörbom, 1989). The modifications consisted of allowing correlations between the by default uncorrelated error terms of different items. Correlations between error terms were theoretically justified by measurement artifacts, like the similar wording in the questionnaire items (Wang & Wang, 2012), and no more than two correlational paths per model were allowed to avoid overfitting.

After finding the best-fitting model for each group, we performed multigroup CFA to test MI. If MI is supported by the data, it indicates that the theoretical construct under study is the same in all groups and measured on the same scale, a prerequisite for meaningful comparisons between groups (Wang & Wang, 2012). To compare factor scores for different groups, Vandenberg and Lance (2000)
and Wang and Wang (2012) recommend that at least three, consecutively more restrictive, levels of MI should be established:

1. Configural invariance: all groups are restricted to have the same number of latent factors and same factor loading patterns (i.e., items load on the same factors in all groups).
2. Metric invariance: equal factor loadings across groups.
3. Scalar invariance: equal factor loadings and item intercepts across groups.

At each new, more restricted, level, we evaluated the fit of the new model using the same GfIs as for the ordinary CFA models. Additionally, we evaluated the deterioration of fit from one level to the next through the change in CFI (ΔCFI). ΔCFI between two consecutively more restricted levels should be smaller than or equal to −0.01 to support invariance between groups (Cheung & Rensvold, 2002). Even if data do not support full configural, metric, or scalar invariance, it is possible to establish partial MI by relaxing the constraints of a few parameters (Byrne, Shavelson, & Muthén, 1989; Wang & Wang, 2012). For example, partial scalar invariance can be supported if invariance demands only are met by allowing at least one intercept to vary freely in one or more groups. According to Dimitrov (2010), up to 20% free intercepts seem to be accepted, though caution and careful judgement in relaxing restraints are recommended (see also Steenkamp & Baumgartner, 1998).

Results

In this section, we first present the results of initial analyses and descriptive statistics. Second, we present the most appropriate CFA models for Sweden and Germany, respectively, and for the separate grades within the two countries. Finally, we present results from measurement invariance testing between grades within the two countries.

Suitability of Data

Going by a recommendation in Tabachnick and Fidell (2013), our data’s suitability for factor analysis was supported by reasonably linear relationships and nothing but significant (at α = .01) correlations between observed variables. Most correlations exceeded .30. Furthermore, the Kaiser–Meyer–Olkin measure of sampling adequacy was .885 in Germany and .902 in Sweden, well above the recommended value of .6 (see Tabachnick & Fidell, 2013). In addition, Bartlett’s test of sphericity was significant, χ²(66) = 13,324.9, p < .001 in Germany and χ²(66) = 8183.2, p < .001 in Sweden.

Descriptive Statistics

As seen in Table 2, Cronbach’s alphas were mostly in the preferred range (as suggested by Nunnally, 1978; Streiner, 2003) between .80 and .90. The most obvious exception was PAv in Germany (.69), but considering the low number of items in each subscale, we deemed the internal consistency for all subscales acceptable.

Subscale means indicated that students in both countries tended to adopt MAp goals more than any performance goals, and PAv goals more than PAp goals. Paired samples t-tests for the Swedish sample confirmed that these differences were significant: MAp and PAp, t (1413) = 23.34, p < .001; MAp and PAv, t (1412) = 20.06, p < .001; and PAp and PAv, t (1418) = 5.79, p < .001. Likewise, for the German sample, there were significant differences between MAp and PAp, t (2653) = 36.81, p < .001; MAp and PAv, t (2654) = 8.66, p < .001; and PAp and PAv, t (2686) = 36.23, p < .001.

Furthermore, a two-sided independent samples t-test showed that the level of PAv goals was significantly higher for German students (M = 3.15, SD = 0.90) than for Swedish students (M = 2.70, SD = 0.99), t (4107) = 14.89, p < .001. There were no other significant differences in goal adoption between the two countries.
**Confirmatory Factor Analysis**

The fit of the five compared models is shown in Table 3. Models C (separating approach from avoidance), D (separating MAP and PAv from PAp), and E (all goals collapsed to one factor) were discarded as the fit for each of these models was worse than for Model A and B in both countries.

For the German students, Model A (three-factor) fitted the data better than Model B (two-factor) and showed acceptable fit on all GFIs except for a TLI value at .874. Modification indices indicated that the fit could be improved by allowing a correlation between the error terms of two items, namely Ag 1 ("In chemistry it is important for me to perform better on tests than the other students") and Ag 2 ("In chemistry my goal is to perform better than other students"). After the modification, the TLI value increased to .905 and all GFIs were on an acceptable level. Our conclusion was that a three-factor model, the revised Model A, described the German students’ achievement goals satisfactorily.

In Sweden, both Model A and B showed acceptable fit to the data as the values for all GFIs were within the acceptable range. Yet, interfactor Pearson correlations showed that PAp and PAv were strongly correlated in Model A ($r = .99$, see Table 4). A Wald test showed that the correlation between PAp and PAv in the Swedish sample was not significantly different from 1 at $\alpha = .05$, $\chi^2(1, n = 1427) = 3.67, p = .06$. This was also true for each individual grade ($p \geq .20$, see Table 4). Thus, although the fit indices of the CFA models indicated that the data fitted both a three-factor and a two-factor model, the high correlations between PAp and PAv indicated that there was no separation of performance goals in approach and avoidance. In addition, the two-factor Model B is more parsimonious and thus preferable.

The PAp–PAv correlations in the unrevised Model A in Germany were lower than in Sweden and significantly different from 1. This was true both for the overall model, including all students, $\chi^2(1, n = 2693) = 188.14, p < .0001$, and when looking at the individual grades (see Table 4). The unrevised

### Table 2. Descriptive statistics and reliabilities for Sweden and Germany.

<table>
<thead>
<tr>
<th>Country</th>
<th>Goal construct</th>
<th>$M$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>MAP</td>
<td>3.25</td>
<td>.86</td>
</tr>
<tr>
<td></td>
<td>PAp</td>
<td>2.59</td>
<td>.85</td>
</tr>
<tr>
<td></td>
<td>PAv</td>
<td>2.70</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td>PAp/PAv</td>
<td>2.65</td>
<td>.90</td>
</tr>
<tr>
<td></td>
<td>MAP/PAp/PAv</td>
<td>2.85</td>
<td>.88</td>
</tr>
<tr>
<td>Germany</td>
<td>MAP</td>
<td>3.47</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td>PAp</td>
<td>2.55</td>
<td>.86</td>
</tr>
<tr>
<td></td>
<td>PAv</td>
<td>3.15</td>
<td>.69</td>
</tr>
<tr>
<td></td>
<td>PAp/PAv</td>
<td>2.84</td>
<td>.85</td>
</tr>
<tr>
<td></td>
<td>MAP/PAp/PAv</td>
<td>3.05</td>
<td>.87</td>
</tr>
</tbody>
</table>

Note: $M =$ unweighted means; $\alpha =$ Cronbach’s alpha; MAP = Mastery-approach; PAp = Performance-approach; PAv = Performance-avoidance.

### Table 3. Goodness-of-fit indicators for achievement goal models in Sweden and Germany.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>TLI</th>
<th>SRMR</th>
<th>$\chi^2$</th>
<th>df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>TLI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>417.7</td>
<td>51</td>
<td>.071</td>
<td>.933</td>
<td>.914</td>
<td>.049</td>
<td>962.2</td>
<td>51</td>
<td>.081</td>
<td>.902</td>
<td>.874</td>
<td>.055</td>
</tr>
<tr>
<td>A*</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>722.1</td>
<td>50</td>
<td>.071</td>
<td>.928</td>
<td>.905</td>
<td>.050</td>
</tr>
<tr>
<td>B</td>
<td>434.7</td>
<td>53</td>
<td>.071</td>
<td>.930</td>
<td>.913</td>
<td>.051</td>
<td>1342.4</td>
<td>53</td>
<td>.095</td>
<td>.862</td>
<td>.828</td>
<td>.063</td>
</tr>
<tr>
<td>C</td>
<td>1697.7</td>
<td>53</td>
<td>.147</td>
<td>.701</td>
<td>.627</td>
<td>.123</td>
<td>2490.5</td>
<td>53</td>
<td>.131</td>
<td>.739</td>
<td>.675</td>
<td>.095</td>
</tr>
<tr>
<td>D</td>
<td>1741.5</td>
<td>54</td>
<td>.148</td>
<td>.693</td>
<td>.624</td>
<td>.124</td>
<td>1888.0</td>
<td>54</td>
<td>.112</td>
<td>.803</td>
<td>.760</td>
<td>.090</td>
</tr>
<tr>
<td>E</td>
<td>1713.7</td>
<td>54</td>
<td>.147</td>
<td>.698</td>
<td>.631</td>
<td>.125</td>
<td>2718.7</td>
<td>54</td>
<td>.135</td>
<td>.714</td>
<td>.651</td>
<td>.095</td>
</tr>
</tbody>
</table>

Note: Fit indices reaching cut-offs are in boldface. df = degrees of freedom, RMSEA = root mean square error of approximation, CFI = comparative fit index, TLI = Tucker-Lewis index, SRMR = root mean square residual. MAP = Mastery-approach; PAp = Performance-approach; PAv = Performance-avoidance. Model A = MAP–PAp–PAv, Model B = MAP–PAp/PAv, Model C = MAP/PAp–PAv, Model D = MAP/PAp/PAv, Model E = MAP/PAp/PAv. $p < .001$ for all chi-square tests.

*Revised version of model A where error terms of item 1 and 2 (both measuring PAp) are allowed to correlate. This model was only fitted for Germany.
Model A was used for calculating the correlations and $p$-values presented in Table 4 as this enabled a fair comparison between the two countries. Similar revisions of Model A in Sweden as in Germany were not possible as interfactor correlations then exceeded 1.0 in most grades, indicating that the models were inadmissible.

Summarizing the results of the CFA models for Sweden and Germany, we find different models appropriate in the two countries. Although a three-factor model fitted German students’ achievement goals, a two-factor model is preferable to describe Swedish students’ achievement goals.

As shown in Table 4, the correlations between PAp and PAv do not follow any increasing or decreasing trend through the grades within any country. However, in the final model for Sweden, the correlation between the MAp and the PAp/PAv factor show an increasing trend over the school years (see Table 5).

Path diagrams for the final models in Sweden and Germany are presented in Figures 2 and 3, respectively. As the figures show, all items had standardized loadings higher than .4 on their respective factors in both models.

When fitting CFA models for each grade within the two countries, the same pattern as for the country level models emerged: Model A described German students’ achievement goals best in all grades but Model B was preferable for Swedish students’ achievement goals in all grades. Still, to attain models with the acceptable fit, post hoc modifications were sometimes necessary in both countries (in Grades 6, 8, 9, and 11 in Sweden and in all grades except for Grade 6 in Germany; see Table 6). In contrast to all other modifications, in Grades 6 and 11 in Sweden, two error term correlations that could not be explained by measurement artifacts were allowed based on modification indices (Ag 2 with Ag 6). The resulting models and their fit statistics are presented in Table 6. With the exception of Grade 5 in Germany, we considered the overall fit of all models acceptable, even though a few models had single GFI s not quite reaching cutoff values. Although the same model fitted all grades within each country, we wanted to make sure that the structure of students’

| Table 4. Correlation between the latent variables performance-approach and performance-avoidance from CFA model A. |
|------------------------|---------------------|---------------------|---------------------|---------------------|
| Grade | Sweden | | Germany | |
| | $r$ | $p^a$ | | $r$ | $p^a$ |
| 5–11 | 0.99 | .06 | .75 | <.001 |
| 5 | >1.00$^b$ | | .76 | <.001 |
| 6 | 0.98 | .68 | .71 | <.001 |
| 7 | 1.00 | .87 | .65 | <.001 |
| 8 | 0.95 | .24 | .75 | <.001 |
| 9 | 0.95 | .20 | .74 | <.001 |
| 10 | 0.98 | .22 | .78 | <.001 |
| 11 | 0.99 | .67 | .72 | <.001 |

Note: All correlations are from the unrevised model A.

*aThe probability of $r$ being equal to 1 (Wald test, $df = 1$).

$bAs this model in inadmissible with $r > 1$, no $p$-value is reported.

| Table 5. Correlation between mastery-approach and performance goals from the best-fitting CFA models of each country. |
|------------------------|---------------------|---------------------|---------------------|---------------------|
| Grade | Sweden, model B | | Germany, revised model A | |
| | MAp–PAp/PAv | | MAp–PAp | MAp–PAv |
| 5–11 | .42 | .79 | .52 | .58 |
| 5 | .25 | .91 | .29 | .61 |
| 6 | .36 | .80 | .47 | .77 |
| 7 | .39 | .70 | .39 | .59 |
| 8 | .45 | .76 | .41 | .31 |
| 9 | .44 | .79 | .57 | .49 |
| 10 | .44 | .80 | .53 | .59 |
| 11 | .48 | .80 | .58 | .54 |

Note: MAp = Mastery-approach; PAp = Performance-approach; PAv = Performance-avoidance.
Measurement Invariance

The results of the MI testing between grades within the two respective countries, using the best model for each grade (shown in Table 6) are shown in Table 7. In both subsamples, GFIs and change in CFI supported configural, metric, and partial scalar invariance across the grades. In conclusion, MI testing indicated that achievement goal models were relatively stable over Grades 5–11 within both Sweden and Germany.

Discussion

Our first research question concerned what achievement goal factor structure is the most appropriate in Sweden and Germany, respectively. We conclude that different goal structures apply for the two countries. In Sweden, a two-goal model comprising a mastery and a performance goal was most
appropriate for describing students’ achievement goals. In Germany, in contrast, a three-goal model splitting performance goals into performance-approach and performance-avoidance goals was most appropriate. This conclusion is based on both the CFA models’ fit statistics and the correlations between factors within the models. CFA showed that a three-goal model had acceptable fit in both Sweden and Germany. However, in Sweden, the three-goal and two-goal CFA model showed an almost equally good fit, but the PAp–PAv correlation was not significantly different from 1 in the three-goal model. Thus, as correlations indicated that PAp and PAv goals were inseparable in Sweden, we deemed the two-goal model to describe the perceptions of achievement goals among Swedish students better. The application of a two-goal model in the Swedish sample stands in contrast to the more generally accepted three-goal model, but is in line with several recent studies that have supported the two-goal model (e.g., Bong et al., 2013; Murphy et al., 2010; Ricco, Schuyten Pierce, & Medinilla, 2010). These results do not imply differences in the levels of goal adoption between the two countries. Despite differences in structure, Swedish and German students may adopt, for example, MAp goals to the same extent. Still, only German students could adopt PAp goals separately from PAv goals as Swedish students do not discriminate between the two.

Though relatively high in both countries, the PAp–PAv correlation was significantly higher in Sweden ($r = .994$) than in Germany ($r = .746$). $z(1427; 2693) = 59.2, p < .001$. High correlations are common, especially when the instrument also measures other than normative aspects of performance goals, for instance, appearance or evaluative aspects (Bong et al., 2013; Hulleman et al.,

---

**Figure 3.** Path diagram showing standardized parameter estimates for the final model (model A) on the German data. Standard errors within parentheses.
Our instrument assessed both the normative and the appearance aspects of performance goals. This could explain the relatively high $\text{PAp}–\text{PAv}$ correlations found in both Sweden and Germany, but not the difference in $\text{PAp}–\text{PAv}$ correlation between the countries. We argue that differences in national culture are a plausible explanation for the difference in $\text{PAp}–\text{PAv}$ correlation between Sweden and Germany observed in this study. There is a striking difference between Sweden and Germany in the competitiveness dimension, offering an intuitively appealing explanation for the difference in $\text{PAp}–\text{PAv}$ correlations. The magnitude of the difference in this dimension does not in itself mean that competitiveness is more strongly linked to the $\text{PAp}–\text{PAv}$ correlation patterns than the other dimensions. Nevertheless, the competitiveness dimension represents a cultural characteristic closely related to performance goals, emphasizing the degree to which people compare themselves with others, which makes it a plausible candidate. Further, the proposed absence of severe consequences of failure in the less competitive Swedish culture (Hofstede 2010; Linnenbrink-Garcia et al., 2012).

<table>
<thead>
<tr>
<th>Grade</th>
<th>$\chi^2$</th>
<th>df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>TLI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden – Model B: $\text{MAp–PAp/PAv}$</td>
<td>5 ($n = 153$)</td>
<td>87.0**</td>
<td>53</td>
<td>.065</td>
<td>.938</td>
<td>.913</td>
</tr>
<tr>
<td></td>
<td>6 ($n = 133$)</td>
<td>92.6***</td>
<td>52</td>
<td>.077</td>
<td>.913</td>
<td>.889</td>
</tr>
<tr>
<td></td>
<td>7 ($n = 280$)</td>
<td>107.1***</td>
<td>53</td>
<td>.060</td>
<td>.950</td>
<td>.938</td>
</tr>
<tr>
<td></td>
<td>8 ($n = 244$)</td>
<td>107.1***</td>
<td>52</td>
<td>.066</td>
<td>.936</td>
<td>.918</td>
</tr>
<tr>
<td></td>
<td>9 ($n = 252$)</td>
<td>131.6***</td>
<td>52</td>
<td>.078</td>
<td>.928</td>
<td>.908</td>
</tr>
<tr>
<td></td>
<td>10 ($n = 212$)</td>
<td>121.7***</td>
<td>53</td>
<td>.078</td>
<td>.939</td>
<td>.924</td>
</tr>
<tr>
<td></td>
<td>11 ($n = 153$)</td>
<td>99.5***</td>
<td>52</td>
<td>.077</td>
<td>.927</td>
<td>.907</td>
</tr>
<tr>
<td>Germany – Model A: $\text{MAp–PAp–PAv}$</td>
<td>5 ($n = 514$)</td>
<td>194.0***</td>
<td>49</td>
<td>.076</td>
<td>.884</td>
<td>.843</td>
</tr>
<tr>
<td></td>
<td>6 ($n = 299$)</td>
<td>113.2***</td>
<td>51</td>
<td>.064</td>
<td>.922</td>
<td>.900</td>
</tr>
<tr>
<td></td>
<td>7 ($n = 272$)</td>
<td>118.8***</td>
<td>50</td>
<td>.071</td>
<td>.919</td>
<td>.893</td>
</tr>
<tr>
<td></td>
<td>8 ($n = 285$)</td>
<td>109.2***</td>
<td>50</td>
<td>.064</td>
<td>.933</td>
<td>.912</td>
</tr>
<tr>
<td></td>
<td>9 ($n = 503$)</td>
<td>197.7***</td>
<td>50</td>
<td>.077</td>
<td>.934</td>
<td>.913</td>
</tr>
<tr>
<td></td>
<td>10 ($n = 341$)</td>
<td>164.324***</td>
<td>49</td>
<td>.083</td>
<td>.927</td>
<td>.902</td>
</tr>
<tr>
<td></td>
<td>11 ($n = 479$)</td>
<td>168.5***</td>
<td>50</td>
<td>.070</td>
<td>.941</td>
<td>.922</td>
</tr>
</tbody>
</table>

Table 6. Goodness-of-fit indicators for final achievement goal CFA models of each individual grade in Sweden and Germany.

Note: Fit indices reaching cut-offs are in boldface. $df$ = degrees of freedom, RMSEA = root mean square error of approximation, CFI = comparative fit index, TLI = Tucker–Lewis index, SRMR = root mean square residual. $\text{MAp}$ = Mastery-approach; $\text{PAp}$ = Performance-approach; $\text{PAv}$ = Performance-avoidance.

<table>
<thead>
<tr>
<th>Grade</th>
<th>$\chi^2$</th>
<th>df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>TLI</th>
<th>SRMR</th>
<th>ΔCFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Con</td>
<td>747.5</td>
<td>367</td>
<td>.071</td>
<td>.935</td>
<td>.918</td>
<td>.062</td>
<td></td>
</tr>
<tr>
<td>Metric</td>
<td>830.8</td>
<td>457</td>
<td>.068</td>
<td>.931</td>
<td>.925</td>
<td>.071</td>
<td>−.004</td>
</tr>
<tr>
<td>Scalar (initial)</td>
<td>954.3</td>
<td>487</td>
<td>.069</td>
<td>.920</td>
<td>.924</td>
<td>.074</td>
<td>−.011</td>
</tr>
<tr>
<td>Scalar (final)</td>
<td>941.4</td>
<td>486</td>
<td>.068</td>
<td>.922</td>
<td>.926</td>
<td>.074</td>
<td>−.009</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Con</td>
<td>1059.0</td>
<td>349</td>
<td>.073</td>
<td>.926</td>
<td>.902</td>
<td>.059</td>
<td></td>
</tr>
<tr>
<td>Metric</td>
<td>1167.5</td>
<td>403</td>
<td>.070</td>
<td>.920</td>
<td>.908</td>
<td>.068</td>
<td>−.006</td>
</tr>
<tr>
<td>Scalar (initial)</td>
<td>1374.3</td>
<td>457</td>
<td>.072</td>
<td>.904</td>
<td>.903</td>
<td>.072</td>
<td>−.016</td>
</tr>
<tr>
<td>Scalar (final)</td>
<td>1304.2</td>
<td>454</td>
<td>.070</td>
<td>.911</td>
<td>.909</td>
<td>.070</td>
<td>−.009</td>
</tr>
</tbody>
</table>

Table 7. Goodness-of-fit indicators for measurement invariance testing between grades within each country.

Note: Fit indices reaching cut-offs are in boldface. $df$ = degrees of freedom, RMSEA = root mean square error of approximation, CFI = comparative fit index, TLI = Tucker–Lewis index, SRMR = root mean square residual. ΔCFI = change in CFI from the invariance level above, i.e., metric compared to configural or scalar compared to metric. $p < .001$ for all chi-square tests.

Analyses are based on the models presented in Table 6.

The intercept of Ag 7 in Grade 9 is freed from restrictions.

The intercepts of Ag 3 and Ag 5 in Grade 5 and Ag 3 in Grade 11 are freed from restrictions.
et al., 2010) could cause the distinction between avoiding failure and approaching success to be less important than how you evaluate what success is. Thus, the normative and appearance subcomponents gain precedence over the valence dimension, making the two performance goals, sharing the normative and appearance subcomponents, inseparable. A similar effect could result from a highly competitive environment, where the consequences of failure are severe. In this case, students may simultaneously strive to succeed and avoid failure, making approach and avoidance goals inseparable, as argued by Bong et al. (2013).

There are contesting explanations for the high correlation between PAp and PAv. For example, Linnenbrink-Garcia et al. (2012) proposed five alternative explanations:

- Age, suggesting higher correlation among younger children than older.
- Perceived competence, suggesting comparatively high correlation for students with low perceived competence.
- Characteristics of the instrument, suggesting comparatively high correlation for instruments with goal relevant PAv items and/or PAp items that to varying extent measure other than the normative subcomponent.
- Specificity, suggesting a higher correlation for general goals than task-specific goals.
- Fear of failure, suggesting comparatively high correlations among individuals high in fear of failure.

Although the objective of our study was not primarily to explain the correlations between achievement goals, we here briefly discuss how these alternative explanations fit our results. The inverse relationship between the age of students and the correlation between PAp and PAv found in earlier studies (e.g., Bong, 2009) was not found in our study. The correlation between PAp and PAv varied somewhat over the grades, but there was no clear trend in any direction. Hence, in line with Hulleman et al. (2010), we found no evidence that pupils in the lower grades would find it more difficult to distinguish between PAp and PAv goals than students in higher grades. Furthermore, we found no indication that the comparatively higher PAp–PAv correlation of the Swedish students was related to lower perceived competence. Perceived competence was not a focus in this study, but was measured through three items (“I’m very talented at chemistry,” “I manage chemical tasks better than my fellow students,” and “I think I am more talented in chemistry than my fellow students”) in the larger study, within which the data for this paper were collected. An independent sample t-test on latent means showed that the perceived competence was significantly lower for German students ($M = -0.03$, $SD = 1.55$) than Swedish students ($M = 0.08$, $SD = 1.35$), $t (2827) = 2.32$, $p = .021$, though the difference in mean was small. Thus, in this study, the student group with slightly higher perceived competence had the highest PAp–PAv correlation, contrary to the hypothesis presented by Linnenbrink-Garcia et al. (2012). As to the proposed influence of the instrument, the same instrument was used in Germany and Sweden to measure achievement goals. Hence, differences in specificity, goal relevance, or focus on normativity in the measurement of performance goals cannot explain the difference in PAp–PAv correlations between the countries. As to the influence of fear of failure on correlations between PAp and PAv, we have not collected any data that allow us to explore this relationship. In sum, none of the four alternative explanations that we were able to examine was applicable on our results.

That the two-goal model was found to be most appropriate for the Swedish sample is in accordance with our previous studies on Swedish students, using partially different instruments and in different cohorts (Grades 5–12, sample size varying between 600 and 5600; Palm, Sullivan Hellgren, & Winberg, 2010; Winberg et al., 2014; Winberg & Palm, 2016). In addition, a recent thesis by Blomgren (2016), to our knowledge the only other Swedish study examining the factor structure of achievement goals, reports the same problem with separating Swedish students’ PAp and PAv goals. The fact that this study replicated our results using different factor analytic approaches, in a different sample, and using a partially different instrument, further supports our conclusion
regarding achievement goals in Sweden. This consistency also supports the hypothesis that culture is an important factor in shaping students’ perceptions of achievement goals. However, although there are theoretical arguments as to why Hofstede et al.’s (2010) cultural dimensions could explain the differences in achievement goal factor structure between Sweden and Germany, our study includes far too few countries to make definite claims regarding the effects of culture. The influence of cultural characteristics on students’ perception of motivation is a highly relevant question for any researcher planning to make cross-cultural comparisons, and for all theorists within motivation research. Hence, multination studies, utilizing validated and multidimensional measures of cultural characteristics, similar to the one by Liu, Jiang, Shalley, Keem, and Zhou (2016), are warranted.

Turning our attention to research question two, concerning the stability of achievement goal models over Grades 5–11, the MI testing indicated that the factor structures were stable in both countries. Thus, comparisons across grades are meaningful within both countries. Yet, in Germany, the fit of model A in Grade 5 was not quite acceptable according to CFI and TLI. It is possible that the younger students had difficulties understanding the questions (see Koskey, Karabenick, Woolley, Bonney, & Dever, 2010), which resulted in a less coherent pattern and thus worse fitting models. However, despite our previous conclusions, it is also possible that this is a result of a less differentiated goal model in lower ages, as suggested by Linnenbrink-Garcia et al. (2012). The correlation between PAp and PAv in Grade 5 was also the highest of all grades in Germany. However, there was no further decreasing trend in the correlation, no similar effect of age was shown in Sweden, and the MI testing showed invariance over the grades, leaving us with very weak evidence of an age effect in Grades 5–11.

Though the PAp–PAv correlation did not show any trend over the grades, the correlations between the MAP and the combined performance goal in Sweden showed an increasing trend. It seems as if, as Swedish students move up through the educational system, goals to learn as much as possible are to an increasing extent paralleled by a desire to compare actual or apparent performance to that of peers. Other studies indicate that having a combination of PAp and MAP, in general, is more productive than having either MAP or PAp goals (see e.g., K. L. Lau & Lee, 2008; Pintrich, 2000b). It is tempting to speculate that our results simply reflect that students adapt to the demands of school. As demands increase with higher grades, students will feel a stronger need to adjust their goals and strategies accordingly. Hence, if students perceive, consciously or not, that a certain combination of goals is more productive, a high proportion of students will adopt this combination of goals, ultimately affecting the correlations between the goals.

**Limitations**

In this study, we have used a composite measure of achievement goals, comprising both the standpoint and standard subcomponents. This limits direct comparisons to studies within the 2 × 2 framework, but should be comparable to studies using instruments based on the trichotomous framework that includes both the standpoint and standard subcomponent.

Pretreatment of the raw data, for example by deleting outliers, could lead to biased results if not performed with caution. In our case, outliers were identified by established methods (see the methods section for a more exhaustive description). Therefore, in total 18% of the responders showing extremely long strings of identical inputs were deleted from the data set. This cleaning leads to slightly lower correlations between achievement goals, but did not substantially affect the differences between Sweden and Germany found in this study. Furthermore, if the effect had been stronger, it would have led to reduced differences between Sweden and Germany, rather than strengthened our result. Hence, we conclude that the cleaning procedure has not affected the conclusions made in this study.

In educational research, data often have a multilevel structure with students clustered in classes, schools, et cetera. If two students from the same cluster are more similar than two students from different clusters, analyses may have to take the clustering into account. In these cases, multilevel
confirmatory factor analysis (MCFA; Muthén, 1994) is preferable over single-level CFA. However, if intraclass correlations (ICCs) are as low as .05, multilevel modeling have few benefits and may be difficult to estimate (Dyer, Hanges, & Hall, 2005). Calculations of ICCs at school level showed that all our items had ICC < .05 in both countries. On the class level, four of the 12 items showed ICC > .05 in Sweden, but only one item was slightly above .1. In Germany, seven of the 12 items showed ICC > .05, but none was above .15. If the multilevel structure is ignored, the possible errors in the estimation of, for example, standard errors, factor correlations, and factor loadings have been shown to be small at these low ICCs, though fit statistics may be impaired (Julian, 2001; Pornprasertmanit, Lee, & Preacher, 2014). In conclusion, taking the multilevel structure of the sample into account should not have altered the results substantially, though the fit of our models may have improved slightly.

The German sample was biased towards students from the higher track of education, the Gymnasium, while no such differentiation was present in the Swedish sample. It is possible that the difference in sample composition affected the outcome of our analyses, and a more balanced German sample would have been preferable. Yet, differences in school performance are not one of the previously mentioned hypothesized explanations to differences in PAp–PAv correlation. School performance may be linked to perceived competence, but we have shown that perceived competence is an unlikely explanation to our results. Indeed, the Swedish students had higher perceived competence than the German, contradicting the intuitive link between higher perceived competence and a sample biased towards higher school performance. Thus, although the effects of the biased sample cannot be completely ruled out, we find it unlikely that it has affected the results in a decisive way.

There are other differences between Swedish and German schools that were not examined in our study but may have influenced our results. Nevertheless, we argue that many of these differences are consequences of the national culture. For example, the fact that children are sorted by previous and expected school achievement in Grade 5 in Germany could be explained by a more competitive culture, where highlighting of differences between individuals may be perceived as less problematic than in the less competitive Sweden. Similar reasoning could be applied on differences in teaching styles, student-teacher relationships, etcetera. Although culture remains as a plausible explanation of the differences in students’ achievement goal factor structures, the design of this study does not allow any certain conclusions on what cultural characteristics are most influential on students’ achievement goals. In the next section, we suggest future studies to answer this question.

Future Directions

To date, there has been little systematic and coordinated research on the effects of culture on students’ achievement goals. The measures of culture, when existing, typically focus on single aspects of national culture, usually the individualistic-collectivistic dimension. King and McInerney (2016) argued that future studies of culture and achievement goals would benefit from exploring other cultural dimensions than the individualistic-collectivistic dimension. We agree with this and argue that, to advance this field of research, future studies should seek to establish what cultural dimensions are most relevant for the level and structure of students’ achievement goals. This research should depart from comprehensive multidimensional descriptions of national culture, such as Hofstede’s Index of National Cultures (Hofstede et al., 2010), and involve a large sample of countries or regions with different cultural profiles. Ideally, this sampling should approximate a factorial design (Lewis-Beck, Bryman, & Liao, 2003), maximizing relevant systematic variation in the cultural dimensions while minimizing the number of observations needed.

Concluding Remarks

This study clearly shows that there were differences between the achievement goal model of Swedish and German students, and it is plausible that these differences are related to differences in the two
countries’ national culture, specifically the level of competitiveness. In Sweden, PAp and PAv goals are indistinguishable and we, therefore, argue for a two-goal model where the two performance goals are merged into one general performance goal. However, we are not arguing for a general revision of the dominant achievement goal models, or to revert to earlier models. Previous studies have shown strong support for the trichotomous model and the $2 \times 2$ model, both including the separation of PAp and PAv. Beside a statistical separation through different factor-analytical techniques, the two performance goal constructs have shown unique antecedents and outcomes in many different samples (Linnenbrink-Garcia et al., 2012). Nevertheless, whether culture is the reason or not, the relation between the goal constructs can apparently vary for different samples in different contexts. Therefore, the results of this study highlight the need to avoid assuming the universality of achievement goals models, and instead acknowledge the role of the specific context of each study. Each time researchers use an instrument in a new context, they should investigate the dimensionality and adjust the model accordingly. Moreover, we like to emphasize the importance of considering other structural aspects than model fit indices when evaluating achievement goal models. In this study, a separation of PAp and PAv in Sweden would have been justified if only GFIs from the CFA were considered. Still, with a correlation not significantly different from 1, such a separation lacks practical significance and should, in our view, be disregarded.

**Acknowledgements**

We thank the German DoLiS-team, especially Ilka Parchmann, Sascha and Andrea Bernholt, and Janet Blankenburg, all at IPN, for cooperation in the data collection and valuable discussions throughout the research process. Thanks also to Ewa Bergqvist and Sofie Areljung at Umeå University for input on earlier versions of the manuscript.

**Disclosure Statement**

No potential conflict of interest was reported by the authors.

**Funding**

This study was financed by the Swedish Research Council (Vetenskapsrådet) [grant number 721-2013-2180].

**ORCID**

Anders Hofverberg [http://orcid.org/0000-0001-8454-319X](http://orcid.org/0000-0001-8454-319X)

Mikael Winberg [http://orcid.org/0000-0002-1535-873X](http://orcid.org/0000-0002-1535-873X)

**References**


