The impact of gender-stereotypical text contents on reading competence in women and men

Kathrin Thums (iD) · Timo Gnambs · Ilka Wolter

Abstract Societies have socially shared assumptions about what constitutes typically male or female attributes. Language can contribute to gender inequality by transmitting gender stereotypes. This study examines whether gender-stereotypical connotations in stimulus texts within a reading competence test might serve as a nuisance factor distorting reading competence measurements. In addition to a general factor for reading competence, we expected gender-stereotypical texts to give rise to gender-specific factors regarding the text content. The research was based on a sample of 813 adults from a pilot study of the *German National Educational Panel Study* (NEPS). A bifactor model confirmed a general factor for reading competence. However, the two gender-specific factors were not found; consequently, no substantial gender differences in reading competence for gender-stereotypical text content were observed. These findings indicate that there is no substantial impact of gender-stereotypical text connotations on the measurement of women's and men's reading competence.

K. Thums (✉)
Leibniz Institute for Educational Trajectories, Wilhelmsplatz 3, 96047 Bamberg, Germany
E-Mail: kathrin.thums@tu-dortmund.de

K. Thums · Prof. Dr. T. Gnambs · Dr. I. Wolter
Institut für Schulentwicklungsforschung (IFS), Technische Universität Dortmund,
Vogelpothsweg 78, 44227 Dortmund, Germany

Prof. Dr. T. Gnambs
E-Mail: timo.gnambs@lifbi.de

Dr. I. Wolter
E-Mail: ilka.wolter@lifbi.de

✷ Springer

Der Einfluss von geschlechtsstereotypen Textinhalten auf die Lesekompetenz von Frauen und Männern

Zusammenfassung Es gibt sozial geteilte Annahmen darüber, was typisch weibliche oder männliche Eigenschaften sind oder sein sollten. Über Sprache werden Stereotype vermittelt, beispielsweise über geschlechtsstereotype Textinhalte, wodurch Ungleichheiten zwischen Frauen und Männern aufrechterhalten werden. Die vorliegende Studie untersucht, ob geschlechtsstereotype Konnotationen von Texten, die zur Messung der Lesekompetenz verwendet werden, einen differenziellen Einfluss für Frauen und Männer auf das Ergebnis der Messung haben. Neben einem generellen Faktor der Lesekompetenz wurde erwartet, dass sich geschlechterspezifische Faktoren identifizieren lassen. Die Stichprobe umfasste 813 Erwachsene aus einer Entwicklungsstudie des Nationalen Bildungspanels (NEPS). Im Bi-Faktormodell konnte der allgemeine Faktor der Lesekompetenz bestätigt werden. Darüber hinaus konnten die beiden geschlechtsspezifischen Faktoren jedoch nicht ermittelt werden. Es zeigten sich keine Unterschiede in der Lesekompetenz hinsichtlich geschlechtsstereotyper Textinhalte. Das Ergebnis deutet darauf hin, dass eine geschlechtsstereotype Textkonnotation keinen substantiellen Einfluss auf die Lesekompetenzmessung von Frauen und Männern hat.

1 Research background and aims

1.1 Reading competence and gender differences in reading competence

A broad range of diverse competences are necessary for active participation in social life. The focus of this study is on reading competence, a fundamental competence for participation in social life among all age groups (OECD 2013b, 2016b). Reading comprehension is described as the product of characteristics of the reader (e.g., gender, prior knowledge, and motivation) and characteristics of texts and tasks (e.g., text content, coherence; cf. Artelt et al. 2001; Christmann and Groeben 1999). One form of this interaction between reader characteristics (gender) and text characteristics (gender-stereotypical text content) was investigated in this study as a potential explanation for gender differences in reading competence between women and men.

The ability to read and understand written language is referred to as reading competence. Reading competence is rooted in cognitive representations and processes and can be learned and trained (Müller and Richter 2014). According to Graesser et al. (1994), reading competence goes beyond the mere decoding of letters, words, and sentences. Rather, this functional and active approach to reading competence emphasizes how a reader engages with the content of a text, obtains information

from the text, and thus draws inferences and places the text content within the global coherence of everyday life (Graesser et al. 1994). Kintsch's (1998) definition of reading competence stresses the interaction between readers' features and text characteristics while emphasizing the active role of the reader, who must integrate prior knowledge and a mental representation of the text during the reading process. In contrast, the current paper mainly relies on the functional *literacy* definition of reading competence, in which the focus is on the resulting product, i.e., the quality of an individual's text representation (Müller and Richter 2014). Consequently, no statements will be made about individual cognitive processes involved in reading.

Reading competence assessments have revealed gender differences throughout school. The results generally suggest that girls achieve higher reading competence scores than boys (e.g., Berendes et al. 2018; Solheim and Lundetræ 2017; Weis et al. 2019), with the largest gender differences in reading during secondary school. In all previous PISA (Programme for International Student Assessment) cycles, 15-year-old girls outperformed boys in reading competence on OECD average (OECD 2019). In contrast, in the PIAAC study (Programme for the International Assessment of Adult Competencies), there were no substantial competence differences between women and men both on OECD average and for the German subsample (OECD 2016a; Solheim and Lundetræ 2017). However, there is much less research on adults' reading competence compared to younger cohorts.

Thus, while previous studies found no substantial gender differences in overall reading competence among adult readers (OECD 2016a), the present study conducts a more differentiated analysis of men's and women's reading competence in interaction with text characteristics, specifically gender-stereotypical text content.

1.2 The role of gender stereotypes in explaining gender differences in reading competence

As previously mentioned, reading comprehension is described as an interaction between readers' characteristics and features of texts or tasks (cf. Artelt et al. 2001; Christmann and Groeben 1999). The present study aimed to further investigate gender differences in reading competence based on a specific form of this proposed interaction between readers' characteristics (i.e., gender) and text characteristics (i.e., gender-stereotypical text content). We expect that men and women process texts with gender-stereotypical text content differently. More precisely, we expect that men have a higher reading competence in texts with stereotypically male connotations than women, who in turn are expected to be superior in reading texts with stereotypically female connotations.

Evidence for this research question can be derived from various research paradigms: The expectancy-value theory (e.g., Eccles 1987; Wigfield et al. 2006), adapted for reading motivation by Möller and Schiefele (2004), provides a potential explanation for gender differences in reading competence. The central components of this theory influencing behavior and motivation, which in turn affect one's competences over the life course (e.g., Becker and McElvany 2018), are expectation beliefs regarding success and failure and subjective value beliefs. According to this theory, women should have lower expectations of success regarding stereotypically

male text contents and consider those contents less important or relevant (i.e., lower value) than men, who in turn should hold lower expectations of success and lower value than women towards stereotypically female content domains.

Individuals' values and expectations are in turn influenced by their social environment. For example, one's social environment includes one's socio-economic status, cultural habits, and experiences with reading, but also gender roles and cultural stereotypes concerning specific academic domains and occupations (Becker and McElvany 2018; Möller and Schiefele 2004). Since gender stereotypes and gender roles are transmitted through the social environment, they are expected to have impact on individuals' interests (values) and self-concept (expectancy) (Möller and Schiefele 2004). Therefore, another possible explanation for gender differences in reading competence is a lower reading-related self-concept—as a consequence of a lower expectation of success—towards stereotypically male text contents in women relative to men, and a lower reading-related self-concept towards stereotypically female text contents in men relative to women. It is plausible to assume that gender stereotypes influence individuals' motivation and interest in specific topics. Since referring a high value to a particular topic leads one to have positive associations with it, which can in turn be an important source of motivation for continuing this activity (Möller and Schiefele 2004), it is important to investigate the impact of gender stereotypes in the context of reading competence.

Gender stereotypes are defined as associating certain characteristics, attitudes, abilities, preferences, and behavior to women and men (Athenstaedt 2003; Prentice and Carranza 2002). Stereotypes are not attributions made about individuals but rather a socially shared assumption regarding the social group in general. Gender stereotypes have both a descriptive (how women or men *typically are*) and a prescriptive component (how women or men *are expected to be*) (Heilman 2012; Prentice and Carranza 2002).

Some of the most common gender stereotypes include that men are (expected to be) agentic, confident, competitive, independent, and assertive (e.g., Diekman and Eagly 2000; Fiske et al. 2002). Other gender stereotypes are that women are (expected to be) affectionate, considerate, warm, and communal (e.g., Diekman and Eagly 2000; Fiske et al. 2002). In addition to these personality traits ascribed to women and men, there are gender stereotypes concerning leisure time activities and occupations (Athenstaedt 2003; Baron et al. 2001; Nosek et al. 2002). Leadership and management positions as well as interest in activities and occupations in STEM fields (science, technology, engineering, and mathematics) are generally perceived as the domains of men (Baron et al. 2001; Nosek et al. 2002). Conversely, social activities, taking care of children, and activities and occupations in the arts, language and social science are gender stereotypes associated with women (Baron et al. 2001; Nosek et al. 2002). Furthermore, research has repeatedly demonstrated achievement-related stereotypes: boys are more interested and achieve higher performance in mathematics (e.g., Nowicki and Lopata 2017; Steffens and Jelenec 2011). However, reading is generally considered a female domain and associated activities are often seen as female-typed (e.g., Nowicki and Lopata 2017; Steffens and Jelenec 2011; Watson et al. 2010). Consequently, there are robust gender differences between women and men with regard to interest in reading and perceived value of reading.

Girls have a higher intrinsic reading motivation and read more frequently on average (Logan and Johnston 2009; McGeown et al. 2012; Schiefele et al. 2012). Likewise, girls' attitudes towards reading are more positive than boys' attitudes (Logan and Johnston 2009).

If these robust and socially shared expectations of women and men (i.e., gender stereotypes) are also reflected in the content of reading material within a reading competence test, gender stereotypes may affect the assessment of men's and women's reading competence. Previous research shows that gender stereotypes are perpetuated and transmitted through formal and semantic features of language in texts as well as text content. Hence, texts' language and content also contribute to enhancing and maintaining gender stereotypes fixed within oral and written language (Garnham et al. 2016).

Various studies have shown that gender differences in overall reading competence can be explained by certain characteristics of the administered instruments. For example, gender differences in reading are affected by the item format: Boys struggle more with open-ended questions compared to girls. Specifically, across several PISA studies, 15-year-old boys experienced more problems with open formats, with their answers containing less or irrelevant information compared to those of girls (e.g., Lafontaine and Monseur 2009; Schwabe et al. 2015). Further studies confirmed that girls have an advantage in continuous text formats and achieve higher reading competence (Artelt and Schlagmüller 2004; Stanat and Kunter 2002).

Studies on the hierarchical lower level of words instead of sentences or texts found qualitative differences between girls and boys in a vocabulary test (McElvany and El-Khechen 2015; McElvany et al. 2016). Girls and boys more often knew words consistent with stereotypes associated with their own gender than words associated with the other gender group.

These results indicate that examining text characteristics might be relevant for explaining gender differences in reading competence. However, no research currently exists on whether the text context, specifically gender-stereotypical content, has an impact on men's and women's reading competence. Therefore, the present study investigated the impact of gender-stereotyped text content in a large-scale assessment of reading competence, extending previous studies by examining differences at the level of longer and more complex texts rather than words or individual sentences (cf. McElvany and El-Khechen 2015; McElvany et al. 2016). The gender-stereotypicality of texts was expected to explain gender differences in reading competence based on the assumption that, in general, women and men have different interests and prior knowledge based on different experiences in their social environments (e.g., education, occupation, leisure time activities; Baron et al. 2001; Christin 2012; Lagaert et al. 2017; Nosek et al. 2002). Women's and men's different levels of interest, previous experiences, and prior knowledge in regard to text contents with gender-stereotypical connotations might initiate that they feel competent and familiar in dealing with different text contents. Consequently, women and men are expected to achieve different competence levels in a reading competence assessment when the text content includes gender stereotypes.

1.3 Research aims and hypotheses

Drawing upon this theoretical and empirical foundation, we hypothesized that gender differences in reading competence can be explained by gender-stereotypical connotations in text content. It is plausible to assume that different levels of prior knowledge and relevant previous experience in different contexts lead to higher familiarity with the text content and therefore make it easier to link information from texts, draw inferences and build coherence with other content. Combined with their aforementioned higher interest in gender-stereotypical domains, this might explain why competence levels in a reading competence test depend on the gender-stereotypical content of the text.

Consequently, men are expected to outperform women in a text whose content addresses, for example, a technical topic, which is a stereotypically male domain (Diekman and Eagly 2000; Heilman 2012). Conversely, women are expected to outperform men on a text with content, for example, about a love story or human relationships, due to the stereotype that women are more emotionally invested in stories and interested in communion (Diekman and Eagly 2000; Heilman 2012).

The hypotheses were as follows:

Hypothesis 1 Alongside a general factor for reading competence, with substantial factor loadings for all tasks on one unidimensional scale, there should be two additional gender-specific factors for reading competence. These gender-specific factors reflect a) stereotypically female text content; and b) stereotypically male text content.

Hypothesis 2 Conditional on the support of hypothesis 1, we expect gender differences in the gender-specific factors for reading competence: a) women have higher reading competence on stereotypically female texts than men, and b) men have higher reading competence on stereotypically male texts than women.

Hence, the hypotheses refer to different levels of analyses. The first hypothesis refers to the material for assessing reading competence, which was tested for further dimensions of reading competence based on gender stereotypes in text content. The second hypothesis refers to interindividual differences (i.e., gender differences) in these specific factors for reading competence. Thus, this second hypothesis relates to the interaction between readers' characteristics (gender) and text characteristics (gender-stereotypical text content).

2 Methods

2.1 Sample and design

The data was collected within a pilot study of the *German National Educational Panel Study* (Blossfeld et al. 2011). The sample was a stratified random sample and predefined using a quota scheme with regard to age and education.

Reading competence was tested among a sample of 939 adults using computer-based assessments in their own households. A subsample of $n = 39$ participants

Table 1 Sample distribution by age and highest level of school education

	N	Percent
Age		
19–25 years	478	58.8
26–45 years	219	26.9
46–65 years	112	13.8
Not specified	4	0.5
Highest level of school education		
Did not complete any form of schooling	2	0.2
Special education school leaving certificate	5	0.6
Basic Hauptschule leaving certificate	57	7.1
Qualifying Hauptschule leaving certificate	36	4.4
Realschule leaving certificate	134	16.5
University of applied sciences entrance qualification (Fachholschulreife)	54	6.6
University entrance qualification (Abitur)	525	64.6

who were assessed in a different test condition as part of the experimental design were excluded. Since only a few participants came from an immigrant background and had a first language other than German ($n = 66$), these participants were also excluded from the analyses. Moreover, all participants born before 1950 ($n = 21$) were excluded from the analyses because we expected different gender attributions and gender stereotypes in older cohorts (Athenstaedt 2000; Eagly et al. 2019; Haines et al. 2016).

Thus, the final sample consisted of 813 adults, of which $n = 424$ were female (52.2%). Their ages ranged from 19 to 65 years old ($M = 30.08$; $SD = 11.16$). A detailed description of the sample distribution in terms of age and school education is displayed in Table 1. Supplementary information about the sample distribution regarding gender, age, and school education is available at: https://osf.io/h8m7a/.

2.2 Research instruments

2.2.1 Reading competence

The reading competence test was based on a consistent framework (Gehrer et al. 2012, 2013) identifying five different text types: informational texts, commentaries or argumentative texts, literary texts, instructional texts, and advertising texts. The pilot study was conducted using a multi-matrix-design with overall 20 text of which participants were randomly assigned to 15 texts and the associated tasks. For this study six informational texts, three literary texts and three advertising texts were included in total in the analyses for the present study. The test consisted of 68 dichotomous multiple-choice tasks as well as more complex polytomous multiple-choice tasks (e.g., ordering tasks, and cloze tasks) related to the content of the respective texts. The tasks posed different cognitive requirements on the reader:

finding information in the text, drawing text-related conclusions (local and global coherence information), and engaging in reflection and assessment.

2.2.2 Background variable gender

Participants indicated their gender in the study questionnaire as either male or female. In the analyses, gender was coded 0 for female and 1 for male.

2.3 Data analysis

2.3.1 Rating and scaling the reading texts

The twelve texts assessing reading competence were categorized a priori with regard to content into texts with a gender-stereotypically male connotation, texts with a gender-neutral connotation and texts with a gender-stereotypically female connotation. This categorization took place purely based on the overall content of each individual text.

The categorization of texts into gender-stereotypically male, gender-stereotypically female and gender-neutral was conducted in two steps. The first step was to derive characteristics from the psychological and educational literature describing stereotypically female and male attributes, characteristics, interests, occupations, and activities based on previous empirical research (Athenstaedt 2003; Athenstaedt et al. 2009; Baron et al. 2001; Christin 2012; Fiske et al. 2002; Haines et al. 2016; Nosek et al. 2002). In the next step, three experts on the topic of gender stereotypes (two of whom are authors of this text) classified the twelve texts with regard to content with gender-stereotypically male, gender-stereotypically female or gender-neutral connotations. Their ratings concerned a text's overall topic on a continuous response scale with five options: very typically male, somewhat typically male, gender-neutral, somewhat typically female, and very typically female. The rating scheme can be found in the appendix. An example of stereotypically female content is a text about yoga and different yoga styles. Yoga is a typically female sport, and more women are interested in and familiar with different yoga traditions then men. An example of stereotypically male content is a text about a photography technique with a lot of technical details about the camera and the underlying technical procedure (Athenstaedt et al. 2009; Christin 2012; Lagaert et al. 2017).

Following the rating procedure, each gender-stereotypical text cluster contained four texts (two informational texts, one literary text and one advertising text). The experts achieved very high levels of consensus according to interrater reliability as measured through the intra-class-correlation, $ICC = 0.86$.

After categorizing of the texts into gender-stereotypical text clusters, the 68 tasks referring to the texts were scaled with a two-parametric logistic item response model (2PL; Birnbaum 1968) using the software package ConQuest (version 4.5.2; Adams et al. 2015). The tasks were scaled separately for each gender-stereotypical text content category. This resulted in three separate scaling models, each on a logit scale. In this way, three well-fitting, unidimensional tests conforming to the 2PL model were developed.

Table 2 Characteristics of gender-stereotyped and neutral texts

Gender-stereotypical female texts	Gender-neutral texts	Gender-stereotypical male texts
1016 words	1162 words	1028 words
23 tasks (13 MC, 10 CMC)	21 tasks (11 MC, 10 CMC)	24 tasks (12 MC, 12 CMC)
Marginal reliability = 0.70	Marginal reliability = 0.66	Marginal reliability = 0.73
Item difficulty $M = -0.97$ ($SD = 1.06$)	Item difficulty $M = -0.18$ ($SD = 0.92$)	Item difficulty $M = -0.75$ ($SD = 1.12$)

MC dichotomous multiple choice, *CMC* polytomous complex multiple choice

Tasks were selected based on their quality (i.e., weighted mean square error, item characteristic curve; Pohl and Carstensen 2013) and discrimination parameters (cf. Rost 2004). Moreover, a differential item functioning (DIF) analysis between men and women was conducted. With the DIF analysis we validated the functions of the single tasks for women and men when controlling for the same reading competence. According to the NEPS scaling guideline, differences between groups of 0.40 to 0.60 represent considerable but not severe cases of DIF, and differences smaller than 0.40 are considered non-considerable DIF (Pohl and Carstensen 2013). All 68 tasks exhibited a DIF below 0.60, which was considered as negligible and the differences were in both directions (i.e., in favor of both genders).

Table 2 compares the three gender-stereotypical text clusters in terms of different indicators after the scaling and DIF analysis. For example, texts in the three categories were comparable in terms of the number of words and tasks. Likewise, the reliability scores were similar and satisfactory for all three categories. The item difficulties for stereotypically male and female texts were very similar. The average item difficulties for tasks with a gender-neutral text connotation was slightly higher ($M = -0.18$, SD = 0.92) than for tasks in the other two text categories (gender-stereotypical female texts $M = -0.97$, $SD = 1.06$ and gender-stereotypical male texts $M = -0.75$, $SD = 1.12$). Particularly, the two gender-stereotypical tests did not differ substantially in their average difficulties; thus, differences in item difficulties are unlikely to bias our analyses (see also footnote on the robustness check on page 8).

2.3.2 Statistical models

A confirmatory ordinal bifactor model was applied to examine the first hypothesis concerning text dimensionality. The bifactor model tested the dimensional structure of the tasks with a latent structural model containing one general factor and two additional group factors. A general factor addresses the target construct (e.g., reading competence) and has a conceptually broader alignment. Therefore, all 68 applied tasks were expected to have substantial loadings ($\beta \geq 0.40$) on the general factor of reading competence. In addition to the general factor, we expected to find group factors representing subdomain constructs (e.g., gender-stereotypical text connotations). Thus, the bifactor model aimed to estimate how much variance is attributable to the general factor and how much to the group factors. The group factors should explain additional specific variance beyond the global factor (Chen et al. 2012; Rodriguez et al. 2016). Because of the large age span in the sample (19 to 65 years old),
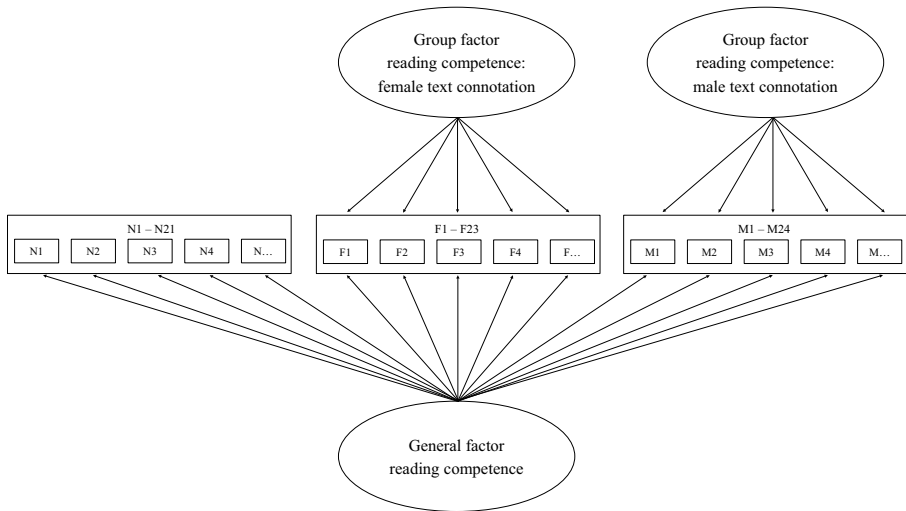
Fig. 1 Hypothesized bifactor model for reading competence with N = tasks with gender-neutral text connotation; F = tasks with female text connotation; M = tasks with male text connotation

we controlled for age differences by regressing all observed indicators and latent variables on the age of the respondents. As a result, the latent factor loadings and variances are corrected for age-related differences. All analyses were also repeated without control variables; but these yielded highly similar results (see supplement).

Fig. 1 illustrates the hypothesized model. Tasks F1 to F23 were expected to have substantial factor loadings on the stereotypically female group factor for reading competence. Tasks M1 to M24 were expected to have factor loadings on the stereotypically male group factor for reading competence. The gender-neutral text connotation (Tasks N1–N21) represented the reference category in this model. Since the factors were set to be orthogonal to each other, there were no correlations among the factors (Reise 2012). The estimator used for the analysis was a maximum likelihood estimation with robust standard errors. The model was implemented in Mplus (version 8; Muthén and Muthén 2017). The Mplus syntax with model input- and output is available at: https://osf.io/h8m7a/.

Afterwards, a multiple group comparison (Kleinke et al. 2017) was computed to test for interindividual differences between women and men in the specific factors for reading competence, as assumed in the second hypothesis. Thus, the average reading competence among women was compared to the average reading competence among men for each factor. In the model, the factor loadings and item intercepts for all reading competence factors were set equal between women and men. The means and standard deviations for the two gender-stereotypical competence factors among women were freely estimated and compared to the means among men. This analysis was also conducted in Mplus (version 8; Muthén and Muthén 2017).

3 Results

3.1 Bifactor model

The results for the bifactor model are summarized in Table 3. The first column shows the standardized loadings for the general factor for reading competence. All 68 tasks loaded significantly ($p < 0.05$) onto the general factor. Most tasks (41 tasks) had noteworthy factor loadings >0.40. The median factor loading for all tasks for the general factor was $Mdn = 0.48$. Reading competence was therefore measured with a general factor.

In contrast, the two gender-stereotypical group factors for reading competence were not confirmed; these findings are illustrated in the second and third columns of Table 3. For both group factors, most loadings were mostly below 0.40. The median factor loading for stereotypically male text group factor was $Mdn = 0.11$, and the median factor loading for the stereotypically female text group factor was $Mdn = 0.29$. For both group factors, there were no substantial loadings or variances beyond the general factor. The results confirmed that a unidimensional model explained the empirical data best.[1]

In addition to the bifactor model results indicating a lack of specific factors for reading competence, we compared the fit of a unidimensional model that included only the general factor (but specific factors) with that of bifactor model. In the bifactor model, the *Akaike information criterion* (AIC) was 48,710 and the *Bayesian information criterion* (BIC) was 50,230. However, in the unidimensional factor model, the AIC was 48,778 and the BIC was 50,018. The AIC showed a better fit for the bifactor model, albeit with a rather small difference (ΔAIC = 8). In contrast, the BIC that also takes the number of estimated parameters into account preferred the unidimensional model. Thus, at least with respect to the BIC, the model fit confirmed the validity of a unidimensional factor model.

To check whether the results depend conditionally on the age of the participants, we examined the model without the age. Similar results were shown and also the unidimensional factor model would be the better solution. All results and the model comparison are available under: https://osf.io/h8m7a/.

Consequently, we cannot confirm Hypothesis 1. In contrast to our hypothesis, there were no additional specific factors for reading competence regarding gender-stereotypical content in the text material.

3.2 Multiple group comparison

To test the second hypothesis at the individual level, a multiple group comparison of the bifactor model was conducted. There were no significant gender differences in reading competences in gender-stereotypical male connotation ($M = -0.19$ logits, $SD = 0.19$, $p = 0.31$; $d = -0.07$).

---

[1] As a robustness check, we reran the model separately for easy (item difficulty < 0) and difficult (item difficulty > 0) tasks in the reading competence test and could not identify specific factors with regard to the gender-stereotypical text content. The output is available at: https://osf.io/h8m7a/.

Table 3 Standardized loading pattern for the bifactor model

Task #	General factor		Group factor for gender-stereotyped female text		Group factor for gender-stereotyped male text	
	B	*S. E.*	*B*	*S. E.*	*B*	*S. E.*
N1	**0.36**	0.06	–	–	–	–
N2	**0.52**	0.05	–	–	–	–
N3	**0.42**	0.07	–	–	–	–
N4	**0.24**	0.05	–	–	–	–
N5	**0.25**	0.05	–	–	–	–
N6	**0.37**	0.05	–	–	–	–
N7	**0.52**	0.05	–	–	–	–
N8	**0.42**	0.06	–	–	–	–
N9	**0.45**	0.06	–	–	–	–
N10	**0.41**	0.05	–	–	–	–
N11	**0.49**	0.05	–	–	–	–
N12	**0.31**	0.06	–	–	–	–
N13	**0.41**	0.05	–	–	–	–
N14	**0.27**	0.05	–	–	–	–
N15	**0.57**	0.04	–	–	–	–
N16	**0.37**	0.05	–	–	–	–
N17	**0.45**	0.05	–	–	–	–
N18	**0.31**	0.05	–	–	–	–
N19	**0.59**	0.03	–	–	–	–
N20	**0.48**	0.04	–	–	–	–
N21	**0.72**	0.03	–	–	–	–
F1	**0.35**	0.06	–0.02	0.22	–	–
F2	**0.53**	0.06	–0.02	0.13	–	–
F3	**0.30**	0.06	0.12	0.21	–	–
F4	**0.48**	0.05	0.16	0.18	–	–
F5	**0.26**	0.06	0.07	0.13	–	–
F6	**0.33**	0.05	–0.03	0.17	–	–
F7	**0.40**	0.05	0.30	0.19	–	–
F8	**0.50**	0.05	0.18	0.17	–	–
F9	**0.41**	0.05	0.20	0.16	–	–
F10	**0.55**	0.05	–0.11	0.19	–	–
F11	**0.53**	0.05	–0.10	0.20	–	–
F12	**0.50**	0.05	–0.15	0.19	–	–
F13	**0.41**	0.06	**0.23**	0.11	–	–
F14	**0.71**	0.05	0.00	0.12	–	–
F15	**0.34**	0.06	–0.10	0.14	–	–
F16	**0.59**	0.06	**0.74**	0.28	–	–
F17	**0.38**	0.07	0.06	0.12	–	–
F18	**0.33**	0.07	**0.52**	0.13	–	–
F19	**0.27**	0.05	–0.13	0.11	–	–
F20	**0.58**	0.04	–0.02	0.11	–	–
F21	**0.51**	0.05	–0.00	0.17	–	–

Table 3 (Continued)

Task #	General factor		Group factor for gender-stereotyped female text		Group factor for gender-stereotyped male text	
	B	*S. E.*	*B*	*S. E.*	*B*	*S. E.*
F22	**0.54**	0.06	0.14	0.16	–	–
F23	**0.43**	0.05	0.02	0.18	–	–
M1	**0.27**	0.05	–	–	0.28	0.15
M2	**0.35**	0.06	–	–	0.08	0.38
M3	**0.30**	0.07	–	–	0.06	0.46
M4	**0.63**	0.06	–	–	0.16	0.30
M5	**0.35**	0.06	–	–	0.26	0.20
M6	**0.37**	0.06	–	–	0.06	0.16
M7	**0.57**	0.05	–	–	0.19	0.16
M8	**0.51**	0.05	–	–	0.16	0.19
M9	**0.37**	0.05	–	–	0.27	0.49
M10	**0.63**	0.05	–	–	0.06	0.44
M11	**0.53**	0.05	–	–	0.02	0.22
M12	**0.51**	0.04	–	–	0.05	0.17
M13	**0.41**	0.05	–	–	0.17	0.22
M14	**0.48**	0.05	–	–	–0.11	0.18
M15	**0.44**	0.06	–	–	0.01	0.28
M16	**0.42**	0.06	–	–	0.22	0.32
M17	**0.46**	0.05	–	–	0.18	0.16
M18	**0.39**	0.05	–	–	0.01	0.29
M19	**0.66**	0.04	–	–	–0.20	0.42
M20	**0.40**	0.05	–	–	–0.10	0.17
M21	**0.65**	0.06	–	–	–0.19	0.51
M22	**0.32**	0.05	–	–	–0.03	0.22
M23	**0.49**	0.06	–	–	0.42	0.87
M24	**0.39**	0.04	–	–	0.04	0.14

N1–N21: tasks with gender-neutral text connotation, reference tasks, *F1–F23*: tasks with female text connotation, *M1–M24*: tasks with male text connotation

B refer to standardized factor loadings; significant loadings are in **bold**, $p < 0.05$

With regard to gender-stereotypical female connotation there was a significant mean difference between women and men (*M = –0.53 logits, SD = 0.21, p < 0.01; d = –0.18*). However, given that only three (F13, F16, and F18) out of 23 tasks showed significant loadings on the specific factor (see Table 3) the observed mean difference did not indicate support for Hypothesis 2. After all, our hypothesis did not refer to individual tasks, but on the gender-stereotypical text connotation.

Therefore, the expected interaction between reader characteristics (gender) and text characteristics (gender-stereotypical text content) in hypothesis 2 was not confirmed.

4 Discussion

4.1 Summary

The main goal of the study was to examine whether gender-stereotypical content in texts making up a reading competence assessment function as an additional dimension of reading competence and explain gender differences in this key competence.

The analyses confirmed a general factor for reading competence at the text level. However, no substantial additional variance was explained by the two hypothesized gender-specific group factors for reading competence. Although prior research has identified stable, socially shared gender stereotypes about typically female and typically male attributions, behaviors and interests (Christin 2012; Diekman and Eagly 2000; Fiske et al. 2002; Haines et al. 2016), which are reflected, for example, in the gender segregation on the labor market (OECD 2013a) and in leisure time activities (e.g., Christin 2012; Lagaert et al. 2017), the findings showed that gender-stereotyped text material in a reading assessment has no impact on reading competence. We assumed on the basis of previous research that gender stereotypes might be activated through text content and thus linked to individuals' prior knowledge and experiences with different topics, increasing their motivation to work with such texts. However, our study could not confirm an effect of gender-stereotypical text content on women's and men's performance in a reading competence assessment.

It is possible that gender stereotypes at the broader level of reading assessment texts have more complex effects on reading competence. McElvany et al. (2016) demonstrated small qualitative gender differences in vocabulary among primary schoolchildren at the hierarchically lower level of words. Children more often knew words consistent with stereotypes associated with their own gender than words associated with the other gender group. Thus, gender stereotypes could be transmitted via different mechanisms, especially in adults, who have a wider range of prior knowledge. Future research could conduct a deeper analysis of text structure—for example, by examining storylines—to disentangle the activation of gender stereotypes through written material (cf. studies regarding grammatical gender representations: Garnham et al. 2012).

Moreover, more complex texts that challenge readers' motivation to work on the reading assessment tasks might be relevant for disentangling the expected effect. Even though our results also found no effect of gender-stereotypical text content when restricted to only difficult items, more complex texts overall might enhance the effect of potential mechanisms for the activation of gender stereotypes (cf. moderators of stereotype threat effect; Martiny and Götz 2011). An experimental setting would provide further opportunities to test similar research hypotheses regarding gender stereotypes in text content. Additionally, it might be worthwhile to examine information about individuals' prior knowledge and familiarity with the text or motivation during the assessment by means of additional questionnaire items or incidental data logged during the computer-based assessment itself. Future research could exploit these data to draw inferences about individuals' process strategies by analyzing jumps between texts and tasks or the time readers spend on texts or tasks.

Concerning the second hypothesis at the individual level, the findings revealed a lack of significant gender differences in the gender-specific factors for reading competence, mostly because the empirical data best fit a unidimensional model with a general latent factor for reading competence only. Women and men were equally competent in reading regardless of potential gender stereotypes in the texts they read. Thus, this study's results were in line with findings from PIAAC, which also reported no substantial competence differences between women and men (OECD 2016a). However, it should be noted that reading abilities as measured in PIAAC do not refer to the same reading comprehension framework as large-scale school-based assessments such as PISA (Lundetræ et al. 2014). Nevertheless, the results from PIAAC and our study indicate that the reading competences of adult women and men are equal and that the gender differences often found in childhood and adolescence seem to decrease over time. A longitudinal analysis of gender differences in reading competence between adolescence and adulthood is necessary to further describe and explain why and how gender differences in this competence domain fade out.

Moreover, a future study could ask a similar question for a younger sample. Since gender differences in reading competence remain constant through adolescence (e.g., Berendes et al. 2018; Solheim and Lundetræ 2017; Weis et al. 2019), it would be interesting to investigate the effect of gender-stereotypical text content in explaining gender differences in reading competence among this age group.

4.2 Limitations

The current study extended previous research on the assessment of reading competence and gender differences in reading competence in a sample of adults. Nevertheless, the study also has some limitations.

First, the participants' age was not representative for the general population of adults in Germany due to an oversampling of university students in this sample. The sample's level of education was also very high on average compared to the German population for this same reason. Research findings indicate that gender stereotypes are related to age and education (e.g., Athenstaedt 2000). It is possible that a younger, more educated sample such as the one in our study may endorse slightly different stereotypes of women and men than a more representative sample regarding age and education, which might have had potential effects on our study findings. However, this means that we might be underestimating the effect in the general population of adults.

Second, the study did not include an explicit measure of whether the texts were perceived as gender-stereotypical by the participants. The current study is based on previous research on gender stereotypes concerning women's and men's experiences, prior knowledge and interest in various domains (e.g., Athenstaedt 2009; Christin 2012; Heilman 2012; Lagaert et al. 2017; OECD 2013a). We applied the same argument to our study by having experts rate the presence of gender stereotypes in the text material. Future research might also add participant ratings to further investigate the individual endorsement of stereotypes and possible mechanisms.

Third, no additional information about participants' occupations, university majors, or leisure time activities were available in this pilot study. This is problematic

because we were not able to include individuals' actual experiences or interests in diverse areas into the model. Collecting data on participants' occupations or leisure time activities may enable future research to control for diverse gender-stereotypical interests or experiences.

5 Conclusion

In conclusion, the current study examined whether gender-stereotypical content in stimulus texts within a reading competence assessment distorts competence measurements and serves as a nuisance factor resulting in gender differences in adults' reading competence.

Gender differences in reading competence have been repeatedly identified. Specifically, adolescent girls are more competent readers than boys (OECD 2019). In adulthood, gender differences seem to vanish (OECD 2016a). Previous findings confirm the presence of consistent and long-lasting gender stereotypes about women and men (e.g., Diekman and Eagly 2000; Haines et al. 2016; Lagaert et al. 2017; OECD 2013a). There is also evidence that language transmits gender-stereotyped attributions and characteristics through representations of words and text processing (e.g., Garnham et al. 2016). However, this study found no evidence that gender-stereotypical text content explains interindividual differences in a reading competence assessment. In light of these results, gender-stereotypical text content appears to be negligible to explain gender differences in adults' reading competence.

References

Adams, R. J., Wu, M. L., & Wilson, M. R. (2015). *ACER ConQuest: Generalised item response modelling software* (Computer software, Version 4). Camberwell: Australian Council for Educational Research.

Artelt, C., & Schlagmüller, M. (2004). Der Umgang mit literarischen Texten als Teilkompetenz im Lesen? Dimensionsanalysen und Ländervergleiche. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Eds.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (pp. 169–196). Wiesbaden: VS.

Artelt, C., Schiefele, U., & Schneider, W. (2001). Predictors of reading literacy. *European Journal of Psychology of Education*, *16*(3), 363–383.

Athenstaedt, U. (2000). Normative Geschlechtsrollenorientierung: Entwicklung und Validierung eines Fragebogens. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *21*(1), 91–104.

Athenstaedt, U. (2003). On the content and structure of the gender role self-concept: Including gender-stereotypical behaviors in addition to traits. *Psychology of Women Quarterly*, *27*(4), 309–318.

Athenstaedt, U., Mikula, G., & Bredt, C. (2009). Gender role self-concept and leisure activities of adolescents. *Sex Roles*, *60*(5–6), 399–409.

Baron, R. A., Markman, G. D., & Hirsa, A. (2001). Perceptions of women and men as entrepreneurs. Evidence for differential effects of attributional augmenting. *Journal of Applied Psychology*, *86*(5), 923–929.

Becker, M., & McElvany, N. (2018). The interplay of gender and social background: A longitudinal study of interaction effects in reading attitudes and behaviour. *The British journal of educational psychology*, *88*(4), 529–549.

Berendes, K., Becker, M., Jacoby, J., Flunger, B., Nagengast, B., & Trautwein, U. (2018). Individuelle Entwicklungsverläufe beim Lesen. Macht das Geschlecht den Unterschied? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *50*(4), 192–208.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, M. R. Novick & A. Birnbaum (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading: Addison-Wesley.

Blossfeld, H.-P., Roßbach, H.-G., & von Maurice, J. (Eds.). (2011). *Education as a lifelong process. The German National Educational Panel Study (NEPS)* (Zeitschrift für Erziehungswissenschaft: 14. Sonderheft). Wiesbaden: VS.

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, *80*(1), 219–251.

Christin, A. (2012). Gender and highbrow cultural participation in the United States. *Poetics*, *40*(5), 423–443.

Christmann, U., & Groeben, N. (1999). Psychologie des Lesens. In B. Franzmann, K. Hasemann, D. Löffler & E. Schön (Eds.), *Handbuch Lesen* (pp. 145–223). München: Saur.

Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, *26*(10), 1171–1188.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed. A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *The American psychologist*, 75(3):301–315. https://doi.org/10.1037/amp0000494.

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly*, *11*(2), 135–172.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902.

Garnham, A., Gabriel, U., Sarrasin, O., Gygax, P., & Oakhill, J. (2012). Gender representation in different languages and grammatical marking on pronouns: When beauticians, musicians, and mechanics remain men. *Discourse Processes*, *49*(6), 481–500.

Garnham, A., Oakhill, J., von Stockhausen, L., & Sczesny, S. (2016). Editorial: Language, cognition, and gender. *Frontiers in Psychology*, *7*, 772.

Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online*, *5*(2), 50–79.

Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S. (2012). *The assessment of reading competence (including sample items for grade 5 and 9). Scientific Use File 2012, Version 1.0.0.* Bamberg. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/1-0-0/com_re_2012_en.pdf. Zugegriffen: 19.06.2020.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395.

Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing ... or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly*, *40*(3), 353–363.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, *32*, 113–135.

Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.

Kleinke, K., Schlüter, E., & Christ, O. (2017). *Strukturgleichungsmodelle mit Mplus. Eine praktische Einführung* (2nd edn.). Berlin: De Gruyter Oldenbourg.

Lafontaine, D., & Monseur, C. (2009). Gender gap in comparative studies of reading comprehension: to what extent do the test characteristics make a difference. *European Educational Research Journal*, *8*(1), 69–79.

Lagaert, S., van Houtte, M., & Roose, H. (2017). Engendering culture: The relationship of gender identity and pressure for gender conformity with adolescents' interests in the arts and literature. *Sex Roles*, *77*(7–8), 482–495.

Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: examining where these differences lie. *Journal of Research in Reading*, *32*(2), 199–214.

Lundetræ, K., Sulkunen, S., Gabrielsen, E., & Malin, A. (2014). A comparison of PIAAC and PISA results. In A. Malin (Ed.), *Associations between age and cognitive foundation skills in the Nordic countries: a closer look at the data* (pp. 171–187). Jyväskylä: University of Jyväskylä, Finnish Institute for Educational Research.

Martiny, S.E., & Götz, T. (2011). Stereotype Threat in Lern- und Leistungssituationen theoretische Ansätze, empirische Befunde und praktische Implikationen. In M. Dresel & L. Lämmle (Eds.), *Motivation, Selbstregulation und Leistungsexzellenz* (Talentförderung, Expertiseentwicklung, Leistungsexzellenz, Vol. 9, pp. 153–178). Münster: LIT.

McElvany, N., & El-Khechen, W. (2015). Mädchenwörter, Jungenwörter – Geschlechtsspezifischer Wortschatz im Grundschulalter? Eine Untersuchung mit Kindern deutscher und anderer Familiensprachen. In H.G. Holtappels, A.S. Willems, M. Pfeifer, W. Bos & N. McElvany (Eds.), *Daten, Beispiele und Perspektiven* (Veröffentlichungen des Instituts für Schulentwicklungsforschung: Jahrbuch der Schulentwicklung, Vol. 18, pp. 270–297). Weinheim: Beltz Juventa.

McElvany, N., El-Khechen, W., Schwabe, F., & Kessels, U. (2016). Qualitative Wortschatzunterschiede zwischen Mädchen und Jungen im Grundschulalter. *Zeitschrift für Pädagogische Psychologie*, *30*(1), 45–55.

McGeown, S.P., Norgate, R., & Warhurst, A. (2012). Exploring intrinsic and extrinsic reading motivation among very good and very poor readers. *Educational Research*, *54*(3), 309–322.

Möller, J., & Schiefele, U. (2004). Motivationale Grundlagen der Lesekompetenz. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Eds.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (pp. 101–124). Wiesbaden: VS.

Müller, B., & Richter, T. (2014). Lesekompetenz. In J. Grabowski (Ed.), *Sinn und Unsinn von Kompetenzen. Fähigkeitskonzepte im Bereich von Sprache, Medien und Kultur* (pp. 29–49). Leverkusen-Opladen: Budrich, Barbara.

Muthén, L.K., & Muthén, B.O. (2017). *Mplus user's guide* (8th edn.). Los Angeles: Muthén & Muthén.

Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002). Math = male, me = female, therefore math not = me. *Journal of Personality and Social Psychology*, *83*(1), 44–59.

Nowicki, E.A., & Lopata, J. (2017). Children's implicit and explicit gender stereotypes about mathematics and reading ability. *Social Psychology of Education*, *20*(2), 329–345.

OECD (2013a). *Closing the gender gap: act now*. Paris: OECD Publishing.

OECD (2013b). *Literacy, numeracy and problem solving in technology-rich environments. Framework for the OECD survey of adult skills*. Paris: OECD Publishing.

OECD (2016a). *Skills matter: Further results from the survey of adult skills*. Paris: OECD Publishing.

OECD (2016b). *PISA 2015 results (Volume 1): excellence and equity in education* (PISA, Vol. 1). Paris: OECD Publishing.

OECD (2019). *PISA 2018 results: what students know and can do*. Paris: OECD Publishing.

Pohl, S., & Carstensen, C.H. (2013). Scaling of competence tests in the National Educational Panel Study—Many questions, some answers, and further challenges. *Journal for Educational Research Online*, *5*(2), 189–216.

Prentice, D.A., & Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: the contents of prescriptive gender stereotypes. *Psychology of Women Quarterly*, *26*(4), 269–281.

Reise, S.P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696.

Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223–237.

Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2nd edn.). Bern: Huber.

Schiefele, U., Schaffner, E., Möller, J., & Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Reading Research Quarterly*, *47*(4), 427–463.

Schwabe, F., McElvany, N., & Trendtel, M. (2015). The school age gender gap in reading achievement: Examining the influences of item format and intrinsic reading motivation. *Reading Research Quarterly*, *50*(2), 219–232.

Solheim, O.J., & Lundetræ, K. (2017). Can test construction account for varying gender differences in international reading achievement tests of children, adolescents and young adults?—A study based on Nordic results in PIRLS, PISA and PIAAC. *Assessment in Education: Principles, Policy & Practice*, *3*, 1–20.

Stanat, P., & Kunter, M. (2002). Geschlechterspezifische Leistungsunterschiede bei Fünfzehnjährigen im internationalen Vergleich. *Zeitschrift für Erziehungswissenschaft*, *4*(1), 28–48.

Steffens, M. C., & Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles*, *64*(5–6), 324–335.

Watson, A., Kehler, M., & Martino, W. (2010). The problem of boys' literacy underachievement. Raising some questions. *Journal of Adolescent & Adult Literacy*, *53*(5), 356–361.

Weis, M., Doroganova, A., Hahnel, C., Becker-Mrotzek, M., Lindauer, T., Artelt, C., & Reiss, K. (2019). Lesekompetenz in PISA 2018 – Ergebnisse in einer digitalen Welt. In K. Reiss, M. Weis, E. Klieme & O. Köller (Eds.), *PISA 2018. Grundbildung im internationalen Vergleich* (1st edition., pp. 47–80). Münster: Waxmann.

Wigfield, A., Eccles, J. S., Roeser, R. W., & Pamela, D.-K. (2006). Development of achievement motivation. In N. Eisenberg & W. Damon (Eds.), *Social, emotional, and personality development* (6th edition Handbook of child psychology, Vol. 3, pp. 933–1002). Hoboken: Wiley.