

Item Wording Effects in Psychological Measures: Do Early Literacy Skills Matter?

Hatice Çiğdem BULUT *

Abstract

While the inclusion of both positively and negatively worded items is a common practice in scales, using positively and negatively worded items together may threaten the validity of a scale. Several studies have been devoted to investigating the effects of item wording methods. The current study investigated item wording effects on the responses of 4028 Turkish fifth-grade students, who responded to the Students Confidence in Mathematics (SCM) and Students Confidence in Science (SCS) scales. The role of early literacy-related variables (i.e., early literacy activities undertaken before primary school, student performance on reading literacy tasks upon entering primary school, and duration of the children's pre-primary school attendance) on item wording effects was also examined. The investigations were conducted using confirmatory factor analysis and the correlated trait–correlated method minus one CFA- CTC(M-1) model, derived from the correlated traits-correlated methods framework. The results indicate that significant item wording effects existed in both scales. Moreover, a significant and positive effect was found in both scales relating to early literacy activities undertaken before school, but no effects were found relating to student performance on reading literacy tasks upon entering primary school or duration of the children's pre-primary school attendance. Overall, the study suggests that researchers and practitioners should consider potential effects when including both positively and negatively worded items in scales, especially scales designed for younger students.

Key Words: Item wording effects, negatively worded items, factor analytic methods, correlated traits-correlated methods, validity.

INTRODUCTION

Educational and psychological scales used in research or large-scale assessments often use a mix of positively and negatively keyed items (e.g., Kam & Meyer, 2015; Michaelides, 2019; Wang, Chen, & Jin, 2015). In the literature, including mixed-format items (i.e., negatively and positively worded items) has been common for a long time (Cronbach, 1950; Nunnally, 1978). In such scales, responses to negatively worded items are routinely recoded to align them with positively worded items so that all items follow the same direction. It is assumed that simply recoding negatively worded items will yield an equivalent opposite measure compared to positively worded items (Marsh, 1996; Nunnally, 1978). However, a considerable amount of research has revealed that negatively worded items might not function as assumed in many cases (e.g., Barnette, 2000; DiStefano & Motl, 2006; Kam & Meyer, 2015). Several studies on the phenomenon of a potential mismatch between intended and interpreted item meanings focus on “item wording effects” as the causal agents (Bolt et al., 2020; Lindwall et al., 2012; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Schmitt & Allik, 2005).

Item wording effects may be related to the respondents' age, race, reading ability, cognitive ability, and/or motivation (e.g., Michaelides, 2019; Schmitt & Allik, 2005; Weems, Onwuegbuzie, & Lustig, 2003; Yang et al., 2012). Many researchers have emphasized the importance of reading ability. In particular, negatively worded items may be more problematic when data is collected from younger respondents due to their level of language and reading skills (Peng et al., 2018). Hence, item wording effects are more likely to occur in large-scale assessments or research focusing upon younger individuals. If self-reporting scales in large-scale assessments are contaminated by variances that are attributable to negatively worded items, this is likely due to a lack of reading comprehension among

* Ph.D., Cukurova University, Faculty of Education, Adana-Turkey, hcayavuz@cu.edu.tr, ORCID ID: 0000-0003-2585-3686

To cite this article:

Bulut, H. C. (2021). Item wording effects in psychological measures: Do early literacy skills matter? *Journal of Measurement and Evaluation in Education and Psychology*, 12(3), 239-253. doi: 10.21031/epod.944067

Received: 28.05.2021

Accepted: 8.09.2021

students in the early grades. These students' interpretation of negatively worded items might lead to inaccuracy in the results, with significant implications relating to derived education policies.

Given the robust relationship between reading ability and early literacy skills, we know that students' early literacy skills contribute to their reading comprehension skills (Lonigan, Burgess, & Anthony, 2000; Storch & Whitehurst, 2002). Therefore, reading practice in early childhood should have a substantial impact on a student becoming a skilled reader (Tunmer & Hoover, 2019). Such practice also might help students to interpret negatively worded items accurately, despite their age. But the effects of early literacy skills in relation to item wording effects have not been deeply researched. We sought to address this gap by examining the relationship between item wording effects and an array of variables related to early literacy activities. We identified potential item wording effects in two different scales applied to fifth-grade students as part of an international, large-scale assessment. Then, we analyzed the relationship between early literacy skills and the discovered item wording effects. We examined whether responses to negatively worded items are different than their counterpart items and whether those responses may have differed due to the early literacy skills of the participants.

Item wording effects

When items in scales include a negative adjective, negative structure, or negative verb conjugation, these items are called "negatively worded items." Self-reporting scales often contain both positively and negatively worded items (e.g., Nunnally, 1978). The reason for this practice is to make respondents more attentive to the content of the items and to avoid response bias (i.e., response styles) in scales (e.g., Barnette, 2000). However, a considerable number of studies have repeatedly shown that including both positively and negatively worded items in a scale might distort factor structure and the inter-item correlation matrix, thereby threatening the validity and reliability of the scale (e.g., DiStefano & Motl, 2006; Kam & Meyer, 2015; Wang et al., 2015). This distortion is thought to be caused by "item wording effects," which refers to artifactual relationships and/or dimensions in a scale caused by the wording of items (Podsakoff et al., 2003).

Item wording effects occur due to the assumption that recoding negatively worded items will guarantee an equivalent opposite measure, equal to positively worded items. For example, let us assume there are two items, "I feel joyful in my school" and "I feel depressed in my school," with two response options, yes or no (this example is inspired by Spector, Van Katwyk, Brannick, and Chen's [1997] work on item direction factors). Considering the related assumption, students who respond yes to the first item should respond no to the second item. However, there might be some students who would say no to both items since those students have more neutral feelings about the school (i.e., feeling neither joyful nor depressed). Such responses could distort the contextualized factor structure of the scale. This example offers a glimpse of how item wording effects occur in scales. There are many other factors (item properties and/or respondents' characteristics) that can also cause item wording effects (Michaelides, 2019; Schmitt & Allik, 2005; Weems et al., 2003; Yang et al., 2012).

Item wording effects can also be related to language and sentence structure (e.g., word order). For example, the dimensionality of the Rosenberg Self-Esteem Scale has been examined in many language families (i.e., Indo-European and Uralic), and different results have been reported (e.g., Lindwall et al., 2012; Pullmann & Allik, 2000). While some languages follow a subject-object-verb (SOV) structure where the subject comes first, the object second, and the verb third, other languages follow a SVO structure (e.g., Turkish). Such linguistic differences play a major role in sentence comprehension (Bornkessel & Schleewsky, 2006) and sentence processing, especially in early language development (Candan et al., 2012). Similarly, sentence negation also varies by sentence structures and language. However, researchers have not considered the relationship between differences in sentence negation and item wording effects.

Item wording effects have been found in scales of self-esteem (e.g., Tomás, Oliver, Galiana, Sancho, and Lila, 2013), anxiety (Weems et al., 2003), perceived stress (Cole, Turner & Gitche, 2019), motivation (Michaelides, 2019), personality (Kam, 2018), and social-emotional learning (Bolt, Wang, Meyer & Pier, 2020). The majority of these studies investigated the occurrence of item wording effects

in the scales using factor analytic methods. However, some of them (e.g., Bolt et al., 2020; Cole et al., 2019; Kam, 2018) utilized different methods to detect item wording effects (e.g., item response theory models or latent difference modeling). On the other hand, some studies investigated which groups of students tend to give inconsistent responses to the negatively worded items (e.g., Kam, 2018; Michaelides, 2019; Weems et al., 2003). These argue that nonalignment between positively and negatively worded items is more likely to occur with younger respondents who possess lower reading abilities or with respondents who seek higher social desirability.

Studies related to reading abilities and item wording effects have emphasized that poor reading ability leads to differential response patterns for positively and negatively worded items in scales (Gnamb & Schroeders, 2020; Weems et al., 2006). Although item wording effects can occur even in samples of graduate students or adolescent participants (Marsh, 1996; Michaelides, 2019; Weems et al., 2006), younger students' reading skills can be more problematic regarding item wording effects due to these participants' lesser development in language acquisition and reading skills (Peng et al., 2018). Michaelides (2019) indicated that the responses of linguistically less proficient respondents led to biased scores obtained from positively and negatively worded items. Given the importance of early literacy skills, as documented by the bulk of extant research (Gustafsson, Hansen, & Rosén, 2013; Melhuish, 2016; Sénéchal & LeFevre, 2002), strong early literacy skills among younger respondents might prevent problems associated with decoding and processing negatively worded items. Some studies show that early literacy skills help to improve students' reading achievement and language skills (e.g., Boyce, Innocenti, Roggman, Norman, & Ortiz, 2010; Gustafsson et al., 2013). Furthermore, these studies have reemphasized that pre-primary education and early literacy skills are very important in the long run. Consistent with this explanation, poor reading ability among younger respondents may be linked to their lesser attainment of early literacy skills. To date, the influence of younger respondents' early literacy skills has not been examined in relation to their processing of negatively worded items. This study builds on previous research that revealed the general importance of reading ability by exploring the specific importance of early literacy skills in item wording interpretation.

Purpose of the Study

This study explores the relationship between item wording effects and literacy activities by asking two research questions (RQs):

RQ 1. Do item wording effects exist in the Students Confidence in Mathematics (SCM) and Students Confidence in Science (SCS) scales?

RQ 2. Is there a relationship between item wording effects and the participants' early literacy skills?

METHOD

Sample

Data were obtained from 4028 Turkish fifth-grade students who participated in the Trends in the International Mathematics and Science Study (TIMSS) 2019 (Mullis, Martin, Foy, Kelly, & Fishbein, 2020). Of the 4028 participants, 1920 were males (47.8% of the sample). In TIMSS, a two-stage random sample design (i.e., firstly schools and then students) is used to select a representative group of students from each country (Mullis & Martin, 2017). TIMSS assesses students' learning outcomes in mathematics and science and provides trends for these subjects. TIMSS also utilizes student, teacher, parent, and school leader questionnaires to gather auxiliary information about the students' home and school contexts (Mullis et al., 2020).

Data Collection Instruments

The Students Confidence in Mathematics (SCM) and Students Confidence in Science (SCS) Scales

In the student questionnaire of TIMSS 2019, there are subject-specific self-reporting scales (i.e., Students Confidence in Mathematics and Students Confidence in Science) due to the strong relationship between the students' academic self-perception and their achievement (Mullis & Martin, 2017). In this study, the SCM and SCS were used to examine item wording effects because both scales include negatively worded items. The SCM consists of nine rating items (five are negatively worded), whereas the SCS consists of seven rating items (four are negatively worded), all measured with a four-point Likert scale (1 = agree a lot, 2 = agree, 3 = disagree, 4 = disagree a lot). Both the SCM and SCS are intended to measure a single underlying latent construct; therefore, an IRT model (i.e., the Rasch partial credit model), based on the unidimensionality assumption, was fitted to the data (Yin & Fishbein, 2020). For the Turkish fifth grade, the alpha reliability coefficients were measured at acceptable levels for the SCM and SCS, at 0.84 and 0.81, respectively (Yin & Fishbein, 2020).

Early literacy-related variables

In the home questionnaire of TIMSS 2019, parents provided information regarding their children's early literacy activities before beginning primary school, their performance on reading literacy tasks upon entering primary school, and the duration of their children's pre-primary school attendance (Mullis & Martin, 2017). In this study, we selected Early Literacy Activities Before School (ASBHELA), Early Literacy Tasks Beginning School (ASBHELT), and Student Attended Preschool (ASDHAPS) as variables. ASBHELA and ASBHELT are index scores calculated by using the Rasch partial credit model (Yin & Fishbein, 2020). The ASBHELA index is derived from items about how often parents performed a set of activities (e.g., reading books, telling stories, writing letters or words) before the child entered school; this was rated with a four-point frequency scale: often, sometimes, never, or almost never. ASBHELT is another index that is derived from items about how well the child performed a set of tasks (e.g., reading some words, reading sentences, reading a story) when the child began the first grade of primary school; this was also measured with a four-point frequency scale: very well, moderately well, not very well, not at all. Lastly, the students' preschool attendance (ASDHAPS) was derived from an item in which parents are asked if and for how long their child attended an early childhood education program; the four-point frequency scale is: 0 = "Did Not Attend" 1 = "1 Year or Less" 2 = "2 Years" 3 = "3 Years or More."

Data Analysis

For the data preparation, first, we recoded positively worded items so that higher scores on all items indicated more positive attributes. Second, the response options of ASDHAPS were combined to create a categorical variable with three levels (i.e., 0 = "Did Not Attend", 1 = "1 Year or Less", and 2 = "2 Years and More"). Then, we checked missing data and confirmed that missing values for each variable were less than 7%.

After the data preparation, the factor structures of the SCM and SCS were evaluated with confirmatory factor analysis (CFA), using Mplus 7 (Muthén & Muthén, 1998–2020). For this, we tested one-factor (Model 1), two-factor (Model 2), and bi-factor models (Model 3). Model 1 hypothesized only one latent factor (i.e., unidimensional model) for each scale as anticipated in the methodology of TIMSS 2019 for SCM and SCS. Model 2 posited two independent latent factors; while one factor was specified for positively worded items, the one factor was specified for negatively worded items. Model 3 assumed one global latent factor and two separate latent factors (i.e., one for the positively worded items and another for the negatively worded items).

To evaluate the presence of item wording effects, we used the correlated traits-correlated methods (CTCM; Marsh, 1989) framework. The CTCM framework is utilized to model multitrait-multimethod (MTMM) data (i.e., data with more than one trait and method). CTCM models enable quantifying the method effects (e.g., item wording effects) by other trait factors and variables so that researchers can

find evidence for method effects with such models (Lindwall et al., 2012). For example, we can specify two method factors (i.e., one for the positively worded items and another for the negatively worded items) in addition to trait factors (i.e., latent factor underlying the items measuring the construct of interest) to examine the validity of a scale (Yang et al., 2012). In the literature, CTCM framework has generally been used to gather convergent and discriminant validity evidence for psychological multi-dimensional constructs (i.e., traits), whose scores were obtained through the different methods. Such models consider the method and trait variance and isolate their variances so that it is possible to model traits without error and method variance (Castro-Schilo, Grimm, & Widaman, 2016).

In this framework, a method factor (i.e., for method effects/item wording effects) can be modeled with negatively worded items. As a result, the trait can be estimated free of the method effects, if there are any. Studies have used CTCM models to investigate methods effects based on negatively worded items (e.g., DiStefano & Motl, 2009; Lindwall et al., 2012; Marsh, 1996; Wu, 2008). However, such models can have convergence and admissibility problems (Fan & Lance, 2017). Therefore, we adapted a correlated trait–correlated method minus one CFA- CTC(M-1) model (Eid, 2000) (Model 4), derived from the CT-CM framework. Eid revised the CFA-CTCM model by specifying the number of method factors (M) minus 1 (e.g., only one method factor is specified either for positively or negatively worded items) to avoid identification problems. Therefore, we modeled only one method factor in this model (Model 4), associated with negatively worded items. Substantive factors (i.e., trait components) and method factors (i.e., factors associated with negatively worded items) are uncorrelated in this model. The difference between the CFA- CTC(M-1) and the CTCM models comes from including only one method but not both factors for positive and negative factors (for details, see Eid, 2000). In the last model (Model 5), we tested the method factor (i.e., item wording effects) with covariates related to early literacy skills. This model predicts the effects of these covariates on substantive factors and method factors. All models are presented in Figure 1.

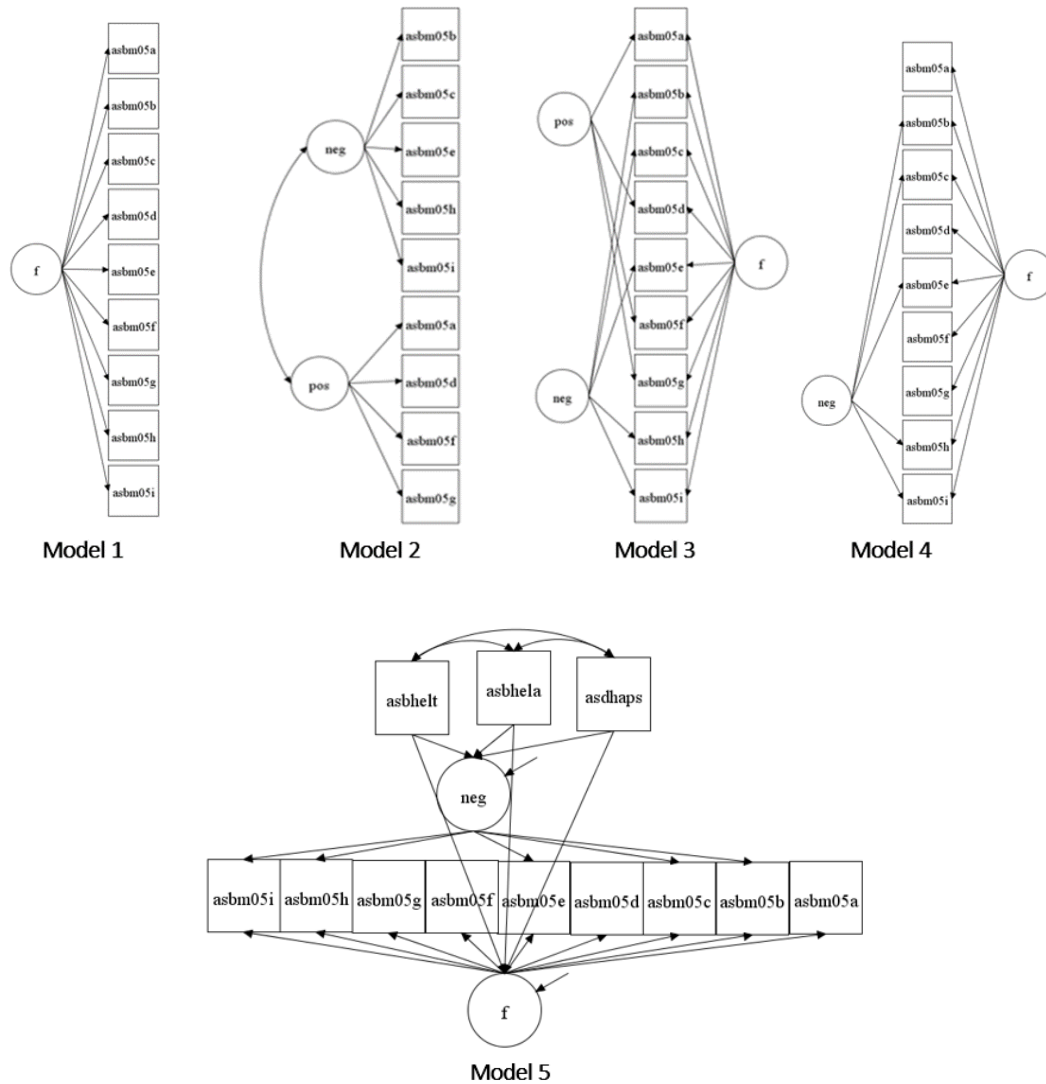


Figure 1. Path Diagrams of the Models (-for SCM)

Note: Pos = Positively worded items; Neg = Negatively worded items; f=Students' confidence in mathematics/science; Asbhelt=Early Literacy Tasks Beginning School; Asbhela= Early Literacy Activities Before School; Asdhaps= Student Attended Preschool.

We used the weighted least square mean and variance adjusted (WLSMV) to estimate the CFA models. To evaluate the models, we used several fit criteria chi-square statistics (χ^2), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the Tucker Lewis Index (TLI). We accepted as a good fit the values of a CFI higher than .95, an RMSEA less than .05, and a TLI higher than .95, based on the recommendations of Hu and Bentler (1999).

RESULTS

Table 1 provides descriptive statistics and item-total correlations for each item. Some negatively worded items had lower mean scores than most of the positively worded items. Item-total correlations ranged from 0.36 to 0.67 ($p < .01$), indicating acceptable discrimination. However, only one item (Item 6 in the SCS = 'My teacher tells me I am good at science') fell outside the criterion (i.e., $< .40$). In addition, the standard deviations of all the negatively worded items were higher than the standard deviations of their counterpart items, indicating high variability within the negatively worded items.

Table 1. Mean, Standard Deviation, Skewness, Kurtosis, and Item-Total Correlations of the Items of the SCR and SCS

Scales	Mean	SD	Skewness	Kurtosis	Item-Total Correlations
SCM					
I usually do well in mathematics/science	3.49	0.75	-1.57	2.31	0.58
Mathematics is more difficult for me than for many of my classmates*	2.91	1.15	-0.44	-1.35	0.60
Mathematics is not one of my strengths*	3.25	1.04	-1.02	-0.41	0.67
I learn things quickly in mathematics	3.36	0.83	-1.28	1.08	0.52
Mathematics makes me nervous*	2.90	1.25	-0.51	-1.44	0.41
I am good at working out difficult mathematics problems	2.94	0.99	-0.65	-0.60	0.53
My teacher tells me I am good at mathematics	3.10	0.97	-0.86	-0.28	0.47
Mathematics is harder for me than any other subject*	2.89	1.18	-0.45	-1.37	0.66
Mathematics makes me confused*	2.79	1.17	-0.30	-1.44	0.61
SCS					
I usually do well in science	3.64	0.66	-2.10	4.64	0.48
Science is more difficult for me than for many of my classmates*	3.12	1.10	-0.77	-0.93	0.57
Science is not one of my strengths*	3.42	0.97	-1.44	0.70	0.65
I learn things quickly in science	3.49	0.79	-1.63	2.17	0.46
My teacher tells me I am good at science	3.19	0.93	-1.03	0.15	0.36
Science is harder for me than any other subject*	3.27	1.06	-1.12	-0.25	0.65
Science makes me confused*	3.16	1.09	-0.88	-0.73	0.60

* Negatively worded items. Source: Mullis, Martin, Foy, Kelly, & Fishbein (2020)

The five models presented in Figure 1 were analyzed for each scale to identify item wording effects. Table 2 presents model chi-square and fit indices for each model. Model 1 represents a one-factor model of a substantive factor (i.e., the SCM or SCS), while Model 2 represents a two-factor model, with two distinct substantive factors (i.e., the negatively worded and positively worded items of the SCM or SCS). Model 3 is a bi-factor model in which there is a general substantive factor underlying all the items and two separate two factors based on the wording of the items. On the other hand, Models 4 and 5 are CTC(M-1) models with a substantive factor (i.e., the SCM or SCS) and a method factor representing negatively worded items. Model 5 specifies the additional effect of three covariates on the method factor and substantive factors. As expected, all the models except Model 1 fit well for the data from both scales. Model 1, which did not consider item wording, provided a poor fit for the data of both scales. For both scales, Model 4 demonstrated a good fit, except for RMSEA, while Model 5 also fit the data well and was slightly better than Model 4. However, the difference between Model 4 and Model 5 is negligible. Overall, these results indicate the presence of item wording effects due to negatively worded items in the SCM and SCS scales.

Table 2. Model fit indexes for the different models for the SCM and SCS scales

SCM	x2	df	RMSEA	CFI	TLI
Model 1	2664.91	27	0.16	0.90	0.88
Model 2	539.79	26	0.07	0.98	0.97
Model 3	182.17	18	0.05	0.99	0.99
Model 4	293.94	22	0.05	0.99	0.98
Model 5	235.01	43	0.03	0.99	0.99
SCS					
Model 1	1762.56	14	0.18	0.92	0.88
Model 2	217.84	13	0.06	0.99	0.98
Model 3	42.42	7	0.03	0.99	0.99
Model 4	142.01	10	0.05	0.99	0.99
Model 5	126.49	25	0.03	0.99	0.99

Table 3 presents the results for the standardized path coefficients of the CTC(M-1) models. In Model 4 for both scales, all parameters were statistically significant, while all negatively worded items' factor loadings were higher for the method factors, except ASBM05C. As for ASBM05E, the factor loading was less than .30 for the substantive factor, while it was higher than .50 for the method factor. In Model 5, ASBHELA had a significant effect on the method factor of both scales ($p < .01$). ASBHELT and ASDHAPS did affect the method factor of the SCS scale ($p < .05$), but measures were nonsignificant for the method factor of the SCM. The size of all the significant effects may be considered low as Model 5 accounted for a low percentage of the variance in the method effects factor, with R^2 values of .03 for both scales.

Table 3. Standardized Path Coefficients for Model 4 and 5

Scales	Model 4		Model 5	
	Substantive factors	Method factors	Substantive factors	Method factors
SCM	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)
ASBM05A	0.87 (0.01) **		0.87 (0.01) **	
ASBM05B	0.52 (0.02) **	0.57 (0.02) **	0.52 (0.02) **	0.56 (0.02) **
ASBM05C	0.66 (0.01) **	0.54 (0.02) **	0.66 (0.01) **	0.53 (0.02) **
ASBM05D	0.78 (0.01) **		0.77 (0.01) **	
ASBM05E	0.29 (0.02) **	0.54 (0.02) **	0.28 (0.02) **	0.53 (0.02) **
ASBM05F	0.78 (0.01) **		0.78 (0.01) **	
ASBM05G	0.71 (0.01) **		0.71 (0.01) **	
ASBM05H	0.57 (0.01) **	0.63 (0.01) **	0.57 (0.02) **	0.64 (0.01) **
ASBM05I	0.50 (0.02) **	0.63 (0.01) **	0.50 (0.02) **	0.64 (0.01) **
ASDHAPS			0.03 (0.02)	0.04 (0.02)
ASBHELA			0.03 (0.03) **	0.14 (0.03) **
ASBHELT			-0.03 (0.02)	0.02 (0.02)
SCS				
ASBS09A	0.86 (0.01) **		0.85 (0.01) **	
ASBS09B	0.44 (0.02) **	0.65 (0.02) **	0.43 (0.02) **	0.66 (0.02) **
ASBS09C	0.60 (0.02) **	0.61 (0.02) **	0.58 (0.02) **	0.62 (0.02) **
ASBS09D	0.79 (0.01) **		0.79 (0.01) **	
ASBS09E	0.65 (0.01) **		0.66 (0.01) **	
ASBS09F	0.50 (0.02) **	0.75 (0.01) **	0.50 (0.02) **	0.75 (0.01) **
ASBS09G	0.50 (0.02) **	0.62 (0.02) **	0.50 (0.02) **	0.63 (0.02) **
ASDHAPS			-0.02 (0.02)	0.06 (0.03) *
ASBHELA			0.08 (0.03) **	0.10 (0.03) **
ASBHELT			0.01 (0.03)	0.07 (0.03) *

** $p < .01$, * $p < .05$. Note: Asbhelt=Early Literacy Tasks Beginning School; Asbhela= Early Literacy Activities Before School; Asdhaps= Student Attended Preschool.

DISCUSSION and CONCLUSION

We examined the role of early literacy-related variables (i.e., early literacy activities undertaken before primary school, student performance on reading literacy tasks upon entering primary school, and duration of the children's pre-primary school attendance) on item wording effects using Turkish fifth graders' responses to the SCM and SCS scales in TIMSS 2019. Both scales were theoretically developed as a unidimensional scale and included negatively worded items. First, we applied several factor-analytic models to identify item wording effects in the scales, and then CFA- CTC(M-1) models to test them with covariates related to early literacy skills. Overall, the findings indicate that the SCM and SCS have item wording effects due to negatively worded items. However, the early literacy-related variables have insignificant or negligible effects and so cannot be used to explain the item wording effects of the SCM and SCS.

Regarding the presence of item wording effects, the results from the CFA models indicate that the inclusion of a second factor underlying the negatively worded items improved the model fit. This

suggests that anticipated factor structures for the SCM and SCS were not maintained in the Turkish sample, which indicates that negatively worded items in the SCM and SCS constituted another factor. Regardless of the subject, obtaining similar results for the confidence scales shows that students answer negatively worded items differently. The result agrees with other conclusions drawn from the literature (e.g., Michaelides, 2019; Wang et al., 2015; Yang et al., 2012). This study shows that students who participate in large-scale assessments display different tendencies when answering items based on their wording. Especially with younger age-group samples, other researchers have shown that negatively worded items might have more deleterious effects (Marsh, 1996; Michaelides, 2019; Weems et al., 2003). This might be due to the younger respondents' reading skills and different interpretations of negatively worded items (e.g., Marsh, 1996; Weems et al., 2003, 2006).

Regarding the second research question, we examined the effects of early literacy-related variables on item wording effects. We found that students' early literacy activities before school entry have significant effects on item wording effects in the SCM and SCS, but low effect sizes were found. Specifically, students engaged in early literacy activities more frequently chose higher response categories in negatively worded items than did students engaged with early literacy activities more frequently. This result indicates that students who had engaged in early literacy activities might more frequently strongly disagree in responses to negative statements compared to moderately agreeing with positively worded items. This is an interesting result and might be related to the students' personality traits (e.g., avoidance motivation, self-consciousness, and neuroticism). Quilty, Oakman, and Risko (2006) state that respondents with higher levels of avoidance motivation or neuroticism are more likely to endorse negatively worded items. Similarly, DiStefano and Molt (2005) found that other personality traits, such as reward responsiveness, fear of negative evaluation, and self-consciousness, contribute to method effects. Therefore, further investigation of the relationships between item wording effects and personality traits across younger age-group samples is recommended as a supplement to the present study. Furthermore, the seemingly counterintuitive findings may be explained by the fact that items related to the variable "students' early literacy activities before school" focus on how often instead of how deeply/successfully students engaged in these early literacy activities. In this case, it can be difficult to decide whether the frequency of doing activities or the success-rate in undertaken activities contributes more to students' early literacy skills.

Students' "performance on reading literacy tasks upon entering primary school" and "years of attending preschool" did not have significant effects on the item wording effects. This result may be due to the students' grade level, as longitudinal studies (e.g., McTigue et al., 2020; Roth, Speece, & Cooper, 2002) indicate that performance differences in early literacy may diminish or the strength of the relationship between achievement and early literacy may decline over the years, due to other sources for variation (e.g., teachers, education quality, and school). This result can additionally be supported by evidence indicating that younger respondents tend to have more problems interpreting the negative expression of a statement (Marsh, 1996; Michaelides, 2019; Weems et al., 2003, 2006). Thus, we conclude that students might interpret negatively worded items differently, regardless of their prior performance or experiences on early literacy activities. Although not a main focus in this study, Model 5 shows insignificant effects of these covariates (i.e., "students' early literacy tasks at the beginning of school" and "years of attending preschool") on the students' self-reported confidence in mathematics and science. Early literacy skills are vital to students' performance in school subjects and attitude development (Caponera, Sestito, & Russo, 2016; Petscher, 2010). However, in this study, students' early literacy skills did not lead to more confident attitudes towards mathematics and science.

Several limitations in this study must be acknowledged. First, we included a limited number of variables related to early literacy skills. Other variables (e.g., letter knowledge, vocabulary, home literacy activities, and family environment) could be included to learn more about the students' early literacy skills. Because TIMSS 2019 did not include these in their parent or student questionnaires, we could not examine the effects of such variables. Second, as data related to early literacy skills were obtained from parents, this can be problematic because self-reported data obtained from parents may be affected by the bias of social desirability. Huang (2017), for example, found that compared with teachers, parents answering the items on behalf of their children are likely to select different response categories depending on children's characteristics (e.g., gender) and parent characteristics (e.g., education level).

Therefore, in our case, parents' responses may also have been affected by these factors. Third, we did not know students' performance ratings related to their literacy skills. As a result, it is unknown whether and how variables related to early literacy skills (e.g., letter knowledge, vocabulary, home literacy activities, and family environment) affected their reading skills and the findings of this study.

Despite these limitations, this study has identified several implications for practice and future research. Firstly, we should take measures to eliminate item wording effects in the scales as much as possible in both the development and administration stages. In the development stage, researchers and practitioners should be careful when including negatively worded phrases, adjectives, and verbs within items. For instance, the item "Mathematics/Science makes me confused" was one of those which had the lowest mean scores in both scales. Therefore, "confused" can be changed to a simpler adjective that is easier for young respondents to interpret. Secondly, given the potential validity threats of item wording effects on scores obtained from scales such as the SCM and SCS, which are used in large-scale assessments, it is important to review negatively worded items in the pilot administration of the scales and to avoid administering scales that include problematic, negatively worded items – especially to relatively younger participants. Thirdly, we recommend that researchers who use data from large-scale assessments check for the presence of item wording effects. If they find evidence for this issue, then it would be beneficial for them to control these effects with a method such as CTCM or the mixed item response theory (IRT) models while estimating scale scores to avoid validity threats. Fourthly, future studies should include students' reading performance and examine how the interactions of reading performance and variables related to early literacy skills affect item wording effects in the scales. Future research also could examine the relationship between reading performance and students' interpretation of negatively worded items using larger and more representative samples and could examine whether the effects of early literacy skills on item wording effects might differ for students in earlier grades.

REFERENCES

- Barnette, J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively- worded stems. *Educational and Psychological Measurement*, 6, 361-370. doi:10.1177/00131640021970592
- Bolt, D., Wang, Y. C., Meyer, R. H., & Pier, L. (2020). An IRT mixture model for rating scale confusion associated with negatively worded items in measures of social-emotional learning. *Applied Measurement in Education*, 33(4), 331–348. doi:10.1080/08957347.2020.1789140
- Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological review*, 113(4), 787. doi: 10.1037/0033-295X.113.4.787
- Boyce, L. K., Innocenti, M. S., Roggman, L. A., Norman, V. K., & Ortiz, E. (2010). Telling stories and making books: Evidence for an intervention to help parents in migrant Head Start families support their children's language and literacy. *Early Education and Development*, 21(3), 343–371. doi:10.1080/10409281003631142
- Candan, A., Küntay, A. C., Yeh, Y. C., Cheung, H., Wagner, L., & Naigles, L. R. (2012). Language and age effects in children's processing of word order. *Cognitive Development*, 27(3), 205-221. doi:10.1016/j.cogdev.2011.12.001
- Caponera, E., Sestito, P., & Russo, P. M. (2016). The influence of reading literacy on mathematics and science achievement. *The Journal of Educational Research*, 109(2), 197-204. doi: 10.1080/00220671.2014.936998
- Castro-Schilo, L., Grimm, K. J., & Widaman, K. F. (2016). Augmenting the Correlated Trait–Correlated Method Model for Multitrait–Multimethod Data. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(6), 798-818. doi:10.1080/10705511.2016.1214919
- Cole, K. L., Turner, R. C., & Gitche, W. D. (2019). A study of polytomous IRT methods and item wording directionality effects on perceived stress items. *Personality and Individual Differences*, 147, 63–72. doi: 10.1016/j.paid.2019.03.046
- Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, 10, 3–31. doi:10.1177/001316445001000101
- DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, 13(3), 440-464. doi: 10.1207/s15328007sem1303_6

- Dodeen, H. (2015). The effects of positively and negatively worded items on the factor structure of the UCLA loneliness scale. *Journal of Psychoeducational Assessment, 33*(3), 259-267. doi: 10.1177/0734282914548325
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika, 65*, 241-261. doi: 10.1007/BF02294377
- Fan, Y., & Lance, C. E. (2017). A reformulated correlated trait–correlated method model for multitrait–multimethod data effectively increases convergence and admissibility rates. *Educational and psychological measurement, 77*(6), 1048-1063. doi:10.1177/0013164416677144
- Gnamb, T., & Schroeders, U. (2020). Cognitive abilities explain wording effects in the Rosenberg Self-Esteem Scale. *Assessment, 27*(2), 404–418. doi:10.1177/1073191117746503
- Gustafsson, J.-E., Hansen, K. Y., & Rosén, M. (2013). Effects of home background on student achievement in reading, mathematics, and science at the fourth grade. In M. O. Martin and I. V. S. Mullis (Eds.), *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade—Implications for early learning*. (pp. 181- 287). Chestnut Hill: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. doi: 10.1080/10705519909540118
- Huang, C. (2017). Cross-informant agreement on the child behavior checklist for youths: A meta-analysis. *Psychological reports, 120*(6), 1096-1116. doi:10.1177/0033294117717733
- Kam, C. C. S. (2018) Novel insights into item keying/valence effect using latent difference (LD) modeling analysis. *Journal of Personality Assessment, 100*(4), 389-397. doi:10.1080/00223891.2017.1369095
- Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods, 18*(3), 512–541. doi:10.1177/1094428115571894
- Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: evidence from a latent-variable longitudinal study. *Developmental psychology, 36*(5), 596. doi:10.1037/0012-1649.36.5.596
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70*, 810–819. doi:10.1037/0022-3514.70.4.810.
- Marsh, H.W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13*, 335-361.
- McTigue, E. M., Schwippert, K., Uppstad, P. H., Lundetræ, K., & Solheim, O. J. (2020). Gender differences in early literacy: boys’ response to formal instruction. *Journal of Educational Psychology*, Advance online publication. doi:10.1037/edu0000626
- Melhuish, E. (2016). Longitudinal research and early years policy development in the UK. *International Journal of Child Care and Education Policy, 10*(1), 1-18. doi:10.1186/s40723-016-0019-1
- Michaelides, M. P. (2019). Negative keying effects in the factor structure of TIMSS 2011 motivation scales and associations with reading achievement. *Applied Measurement in Education, 32*(4), 365–378. doi: 10.1080/08957347.2019.1660349
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/timss2019/frameworks/>
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/international-results/>
- Muthén, L. K., & Muthén, B. O. (1998-2020). *Mplus user’s guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Nunnally, J.C. (1978). *Psychometric Theory* (2nd edn). McGraw-Hill.
- Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., ... & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological bulletin, 144*(1), 48. doi:10.1037/bul0000124
- Petscher, Y. (2010). A meta-analysis of the relationship between student attitudes towards reading and achievement in reading. *Journal of research in reading, 33*(4), 335-355. doi:10.1111/j.1467-9817.2009.01418.x
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. doi:10.1037/0021-9010.88.5.879
- Pullmann, H., & Allik, J. (2000). The Rosenberg Self-Esteem Scale: its dimensionality, stability and personality correlates in Estonian. *Personality and Individual Differences, 28*(4), 701-715. doi: 10.1016/S0191-8869(99)00132-4

- Quilty, L. C., Oakman, J. M., & Risko, E. (2006). Correlates of the Rosenberg self-esteem scale method effects. *Structural Equation Modeling*, 13(1), 99-117. doi:10.1207/s15328007sem1301_5
- R Development Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *The Journal of Educational Research*, 95(5), 259-272. doi: 10.1080/00220670209596600
- Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg self-esteem scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology*, 89, 623-642. doi:10.1037/0022-3514.89.4.623
- Sénéchal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child development*, 73(2), 445-460. doi:10.1111/1467-8624.00417
- Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management*, 23(5), 659-677.
- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental psychology*, 38(6), 934. doi:10.1037/0012-1649.38.6.934
- Tomás, J. M., Oliver, A., Galiana, L., Sancho, P., & Lila, M. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(2), 299-313. doi:10.1080/10705511.2013.769394
- Tunmer, W. E., & Hoover, W. A. (2019). The cognitive foundations of learning to read: A framework for preventing and remediating reading difficulties. *Australian Journal of Learning Difficulties*, 24(1), 75-93. doi:10.1080/19404158.2019.1614081
- Wang, W., Chen, H., & Jin, K. (2015). Item response theory models for wording effects in mixed-format scales. *Educational and Psychological Measurement*, 75(1), 157-178. doi:10.1177/0013164414528209
- Weems, G. H., Onwuegbuzie, A. J., & Collins, K. M. (2006). The role of reading comprehension in responses to positively and negatively worded items on rating scales. *Evaluation & Research in Education*, 19(1), 3-20. doi:10.1080/09500790608668322
- Weems, G. H., Onwuegbuzie, A. J., & Lustig, D. (2003). Profiles of respondents who respond inconsistently to positively- and negatively-worded items on rating scales. *Evaluation & Research in Education*, 17(1), 45-60. doi:10.1080/14664200308668290
- Yang, Y., Chen, Y. H., Lo, W. J., & Turner, J. E. (2012). Cross-cultural evaluation of item wording effects on an attitudinal scale. *Journal of Psychoeducational Assessment*, 30(5), 509-519. doi: 10.1177/0734282911435461

Psikolojik Ölçeklerde Madde İfade Etkisi: Erken Okuryazarlık Becerileri Fark Yaratıyor Mu?

Giriş

Eğitim ve psikoloji alanında kullanılan ölçeklerde olumlu ve olumsuz yönde ifade edilmiş maddeler birlikte bulunabilmektedir (örn., Kam & Meyer, 2015; Michaelides, 2019; Wang ve diğ., 2015). Bunun nedeni olarak bu tür ölçme araçlarında olumlu yönde ifade edilmiş maddelerin yanında olumsuz yönde ifade edilmiş maddelerin yer almasının yaygın bir yaklaşım olması gösterilebilir (Cronbach, 1950; Nunnally, 1978). Bu tür ölçme araçlarında olumsuz yönde ifade edilmiş maddeler ters kodlanarak puanlamaya katılır. Bu işlemle birlikte bu maddelerin olumlu yönde ifade edilmiş maddeler gibi çalışacağı varsayılmaktadır (Marsh, 1996; Nunnally, 1978). Fakat, alan yazındaki çalışmalar olumsuz yönde ifade edilmiş maddelerin varsayılan şekilde işlemediğini ortaya koymaktadır (örn., Barnette, 2000; DiStefano & Motl, 2006; Kam & Meyer, 2015). Yapılan bazı araştırmalar, olumsuz yönde ifade edilen maddelerin ölçme aracından elde edilen puanların geçerliğini tehdit ettiğini ve güvenilirliğini düşürdüğünü göstermektedir (Barnette, 2000; DiStefano & Motl, 2006; Kam & Meyer, 2015). Bunun nedeni olarak ise alan yazında madde ifade etkisi (item wording effect) olarak tanımlanan, bireylerin olumsuz yönde ifade edilmiş maddeleri farklı anlamlandırmalarından dolayı olumsuz maddelerin kendi aralarında ayrı bir faktör oluşturması durumu görülmektedir (DiStefano & Motl, 2006; Dodeen, 2015).

Madde ifade etkisi, olumsuz maddelerin ters kodlandıktan sonra olumlu ifade edilmiş maddeler gibi aynı yönde ve ölçme gücünde işlemediğinde ortaya çıkmaktadır. Bu duruma bir örnek verecek olursak, bir ölçekte şu iki maddenin olduğunu düşünelim: “Okulumda kendimi neşeli hissediyorum.” ve “Okulumda kendimi depresif hissediyorum.” (Örnek Spector ve diğerlerinin [1997] çalışmasından uyarlanmıştır). İlgili varsayım düşünüldüğünde ilk maddeye evet yanıtını veren bireyin ikinci maddeye hayır yanıtını vermesi beklenir. İkinci maddeye verilen yanıtlar ters kodlandığında birinci maddeye benzer şekilde yanıt örüntülerinin oluşacağı varsayılır. Fakat bazı yanıtlayıcılar okullarında kendilerini ne neşeli ne de depresif hissetmedikleri için her iki maddeye de hayır yanıtını verebilir. Bu durumda, ikinci madde ters kodlandığında veri setinde öngörülme yanıt örüntüleri ortaya çıkmaktadır. Bu tür yanıt veren bireylerin yanıtları veri setinde olduğunda ilgili ölçeğin teorik açıdan öngörülen faktör yapısı etkilenmektedir. Bu basit örnek sadece madde ifade etkisinin nasıl oluşabileceğini anlatmak için verilmiştir. Bunun yanında, madde ifade etkisinin oluşmasına yol açan birçok değişken (madde ve/veya yanıtlayıcı özellikleri) bulunmaktadır (Michaelides, 2019; Schmitt & Allik, 2005; Weems ve diğ., 2003; Yang ve diğ., 2012).

Alan yazındaki çalışmalar, ölçeklerde madde ifade etkisinin yanıtlayıcıların yaşına, kültürel özelliklerine, okuduğunu anlama becerilerine, bilişsel becerilerine ve motivasyonlarına göre ortaya çıkabileceğini göstermişlerdir (örn., Michaelides, 2019; Schmitt & Allik, 2005; Weems ve diğ., 2003; Yang ve diğ., 2012). Bu konuda yapılan çalışmalar okuduğunu anlamamanın önemini vurgulamaktadır. Özellikle, olumsuz yönde ifade edilmiş maddelerin küçük yaş gruplarına uygulanan ölçeklerde daha fazla problem yarattığı belirtilmektedir. Bunun nedeni olarak bu yaş grubundaki bireylerin dil ve okuduğunu anlama becerilerinin hala gelişim sürecinde olması gösterilmektedir (Peng ve diğ., 2018).

Okuduğunu anlama becerileri ile erken okuryazarlık becerileri arasındaki ilişki dikkate alındığında, öğrencilerin erken okuryazarlık becerilerinin okuduğunu anlama becerilerinde önemli bir rol oynadığını bilinmektedir (Lonigan ve diğ., 2000; Storch & Whitehurst, 2002). Bu nedenle, erken çocukluk döneminde okuma aktiviteleriyle ilgili daha çok tecrübe sahibi olan bireyler okuduklarını daha iyi anlamaktadırlar (Tunmer & Hoover, 2019). Buradan hareketle, bu bireylerin yaşlarına rağmen olumsuz yönde ifade edilmiş maddeleri doğru şekilde anlamlandırması beklenebilir. Bahsedilen ilişkinin önemine rağmen, alan yazında erken okuryazarlık becerilerinin olumsuz madde etkisinde bir etkisi olup olmadığı çalışılmamıştır. Bu nedenle, bu çalışmanın amacı erken okuryazarlık becerileriyle ilgili olan aktivitelerin olumsuz yönde ifade edilmiş maddeleri anlamlandırmada farklılık yaratıp yaratmadığını incelemektir. Bunun için beşinci sınıf öğrencilerine uygulanmış geniş ölçekli bir testte yer alan iki farklı ölçekte madde ifade etkisinin varlığı araştırılmıştır. Bunun yanında, bazı erken okuryazarlıkla ilgili değişkenlerin olası bu etki üzerindeki rolü incelenmiştir.

Yöntem

Bu çalışmanın örneklemini Uluslararası Matematik ve Fen Eğilimleri Araştırması (the Trends in International Mathematics and Science Study [TIMSS]) 2019’a katılmış 4028 (%47.8 erkek) beşinci sınıf Türk öğrencileri oluşturmaktadır (Mullis ve diğ., 2020). TIMSS, dört yılda bir katılımcı ülkelerin dördüncü/ beşinci ve sekizinci sınıf öğrencilerinin matematik ve fen alanlarında başarılarını belirlemeyi amaçlamaktadır. Ayrıca, TIMSS öğrencilerden, öğrencilerin öğretmenlerinden ve okul yöneticilerinden çok yönlü bilgi toplamaktadır. Bu amaçla, TIMSS başarı testleri dışında birçok ölçeği de içinde bulunduran anketleri de uygulanmaktadır.

Bu çalışmada, dörtlü Likert tipinde olan “Matematik Dersinde Kendine Güvenme” (The Students Confident in Mathematics [SCM]) ve “Fen Bilimleri Dersinde Kendine Güvenme” (The Students Confident in Science [SCS]) ölçekleri kullanılmıştır (Mullis ve diğ., 2020). SCM’de beş olumsuz ve dört olumlu yönde ifade edilmiş madde, SCS’de ise dört olumsuz ve üç olumlu yönde ifade edilmiş madde bulunmaktadır. Ölçeklerin teorik açıdan tek boyutlu olduğu ifade edilmektedir. Türk öğrencilerinin veri setlerinde, alfa güvenirlik katsayıları SCM ve SCS için sırasıyla 0.84 ve 0.81 olduğundan, ölçeklerin güvenirlik katsayıları kabul edilebilir düzeydedir (Yin & Fishbein, 2020).

TIMSS 2019’da ev anketinde ebeveynler çocuklarının erken okuryazarlıklarına ilişkin bazı soruları yanıtlamışlardır. Bu çalışmada, bu anketten “Okuldan Önce Yapılan Erken Okuryazarlık Aktiviteleri”

(ASBHELA), “Okula Başlarken Yapılan Erken Okuryazarlık Çalışmaları” (ASBHELT) ve “Öğrencilerin Okul Öncesi Eğitime Katılımı” (ASDHAPS) değişkenleri ele alınmıştır. ASBHELA ve ASBHELT Rasch kısmi puanlama modeli kullanılarak hesaplanan indeks puanlarıdır (Yin & Fishbein, 2020). ASDHAPS ise öğrencilerin okul öncesi eğitime katılıp katılmadığını, katıldıysa ne kadar katıldığını gösteren kategorik bir değişkendir.

Çalışmada verilerin analizinde öncelikle kayıp veriler incelenmiştir ve her bir değişkenin kayıp veri değerinin %7’den az olduğu görülmüştür. Sonrasında olumsuz maddeler ters kodlanmıştır. Bu amaçla öncelikle SCM ve SCS’nin faktör yapısı tek -faktör (Model 1), iki-faktör (Model 2) ve bifaktör modeli (Model 3) ile incelenmiştir. Madde ifade etkisinin varlığı ise ilişkili- özellik ilişkili yöntem (correlated traits-correlated methods-[CTCM; Marsh, 1989]) modeli kullanılarak incelenmiştir. CTCM, çoklu özellik-çoklu yöntem matrislerini modellemede kullanılmaktadır. Bu model çerçevesinde bir yöntem faktörü (yöntem etkisi/ madde ifade etkisi) modele dahil edilerek, özellikler/ gizil yapılar (traits) bu yöntemin etkisi kaldırılarak kestirilebilir. Bunun yanında, bu tür modeller yakınsama ve kabul edilebilirlik problemleri gösterebilmektedir (Fan & Lance, 2017). Bu nedenle, çalışmada ilişkili özellik- ilişkili yöntem (M-1) modeli kullanılmıştır (correlated trait-correlated method minus one CFA-CTC(M-1) model [Eid, 2000]) (Model 4). Bu modelde, olumsuz yönde ifade edilmiş maddelerin bağlandığı sadece bir yöntem faktörü tanımlanmıştır. Bu yöntem faktörü ile gizil değişkene ilişkin faktörler arasındaki korelasyon tanımlanmamıştır. Son modelde (Model 5), Model 4’te tanımlanan yöntem faktörüne ve gizil değişkene ilişkin faktöre erken okuryazarlıkla ilgili üç kovaryant değişkeni eklenmiştir. Modellerin değerlendirilmesinde ki-kare istatistiği (χ^2) ve bazı uyum indeksleri (Tucker Lewis indeksi - the Tucker Lewis Index [TLI], Karşılaştırmalı uyum indeksi- Comparative Fit Index [CFI], Ortalama hata karekök yaklaşımı- Root mean square error approximation [RMSEA]) dikkate alınmıştır. Modellerin performans kriteri olarak RMSEA’nın .05’ten düşük olması, TLI ve CFI’nın ise .95’ten büyük olması dikkate alınmıştır. Analizlerin hepsi Mplus 7 (Muthén & Muthén, 1998–2020) ve R (R Development Core Team, 2021) kullanılarak yapılmıştır.

Sonuç ve Tartışma

Bu çalışmada, teorik açıdan tek boyutlu olduğu öngörülen, beşinci sınıf Türk öğrencilere uygulanmış SCM ve SCS ölçeklerinde madde ifade etkisinin olup olmadığı incelenmiştir. Aynı zamanda, bu çalışmada erken okuryazarlık becerileriyle ilgili olan aktivitelerin, olumsuz yönde ifade edilmiş maddelerin anlamlandırmasında farklılık yaratıp yaratmadığı da araştırılmıştır. Çalışmada analiz edilen Model 1, 2, 3 ve 4’ün sonuçları ele alındığında hem SCM hem de SCR’de madde ifade etkisinin olduğu belirlenmiştir. Olumsuz yönde ifade edilmiş maddeler için ayrı tanımlanmış faktörün olduğu modeller daha iyi uyum göstermiştir. Özet olarak, öğrencilerin maddeleri ifade edilmiş yönlerine göre farklı yorumladıklarını belirtilebilir. Bu durum, alinyazındaki birçok çalışma ile paralellik göstermektedir (Michaelides, 2019; Wang ve diğ., 2015; Yang ve diğ., 2012). Araştırmalar özellikle küçük yaş gruplarında madde ifade etkisinin daha etkili olabileceğini belirtmektedir (Marsh, 1996; Michaelides, 2019; Weems ve diğ., 2003). Bunun nedeni, özellikle yaşı küçük bireylerin olumsuz yönde ifade edilmiş maddeleri anlamlandırmada daha fazla zorluk yaşaması olarak gösterilmektedir.

Çalışmada Model 5’in sonuçlarına göre öğrencilerin erken okuryazarlık aktiviteleri değişkeniyle SCM ve SCS’de bulunan madde ifade etkisi arasında manidar ve pozitif ilişki bulunmaktadır. Bu ilişkinin etki büyüklüğü ise düşük düzeydedir. Buna göre bu öğrenciler olumsuz yönde ifade edilmiş maddeleri yanıtlarken daha yüksek kategorileri tercih etmektedirler. Bunun yanında, çalışmadaki diğer değişkenler ile madde ifade etkisi arasında manidar ilişki bulunmamıştır. Tüm bu bulguların nedeni olarak erken okuryazarlıkla ilgili becerilerin okuduğunu anlamaya etkisinin zamanla azalıyor olması belirtilebilir. Bu konuda yapılan boylamsal çalışmalar bu durumu desteklemektedir (McTigue ve diğ., 2020; Roth ve diğ., 2002).

Bu çalışmanın bulguları, sınırlıklar çerçevesinde değerlendirilmelidir. Öncelikle, çalışmaya erken okuryazarlıkla ilgili sınırlı sayıda değişken dahil edilmiştir. Aynı zamanda dahil edilen değişkenler, öğrenciler ya da öğretmenler yerine ebeveynler tarafından yanıtlandırılan ölçme aracından elde edilmiştir. Alan yazındaki çalışmalar, ebeveynlerin ölçeklere verdikleri yanıtların sosyal beğenirlikten

etkilenileceğini göstermektedir (Huang, 2017). Son olarak, bu çalışmada öğrencilerin okuduğunu anlama becerilerine ilişkin başarı puanları bulunmamaktaydı. Bu nedenle, erken okuryazarlık becerileri yüksek düzeyde olan öğrencilerin okuduğunu anlamada ne kadar başarılı olduğu bilinmemekteydi.

Çalışmanın sınırlılıklarına rağmen araştırmacılara ve uygulayıcılara bazı öneriler sunulabilir. Birinci olarak, ölçek uyarlama ya da geliştirme çalışmalarında, özellikle uygulama yapılacak yaş grubu dikkate alınarak olumsuz yönde ifade edilen maddelerin incelenmesi önerilir. Ayrıca bu tür maddelerin öngörülen faktör yapısını tehdit edip etmediği de araştırılmalıdır. Eğer araştırmacılar ya da uygulayıcılar hali-hazırda kullanılan ve olumsuz yönde ifade edilmiş maddeler içeren ölçekleri kullanacaklarsa, bu ölçeklerde madde ifade etkisinin varlığını kontrol etmeleri uygun olacaktır. Eğer ölçeklerde madde ifade etkisi varsa alan yazında önerilen yöntemlerle madde ifade etkisi kaldırılarak yanıtlayıcıların ölçek puanları bu şekilde hesaplanmalıdır.