

Received: April 19, 2015

Revision received: January 4, 2016

Accepted: December 14, 2016

OnlineFirst: April 10, 2017

Copyright © 2017 EDAM

www.estp.com.tr

DOI 10.12738/estp.2017.3.0051 • June 2017 • 17(3) • 721–744

Research Article

An Evaluation of the Psychometric Properties of Three Different Forms of Daly and Miller's Writing Apprehension Test through Rasch Analysis

Neşe Güler¹
Sakarya University

Ahmet Güneyl³
Near East University

Mustafa İlhan²
Dicle University

Süleyman Demir⁴
Sakarya University

Abstract

This study evaluates the psychometric properties of three different forms of the Writing Apprehension Test (WAT; Daly & Miller, 1975) through Rasch analysis. For this purpose, the fit statistics and correlation coefficients, and the reliability, separation ratio, and chi-square values for the facets of item and person calculated for the 26-item, one-dimensional, 21-item, one-dimensional and 21-item, four-dimensional forms of the test were compared. The study was conducted with 720 secondary-school students in Nicosia, Northern Cyprus. Having excluded the incomplete or incorrectly completed measurement tools, data for 604 students remained in the data set. The data obtained from the research were analyzed through the Rasch model by making use of the FACETS package program. The results demonstrated that the 21-item, one-dimensional model was the most appropriate model for the WAT. Aside from this, more accurate estimations were found able to be made for students' writing apprehension by adding items with differing levels of difficulty (higher or lower than existing items) into the test.

Keywords

Writing apprehension • Writing Apprehension Test • The Rasch model • Category statistics

¹ Department of Measurement and Evaluation in Education, Faculty of Education, Sakarya University, Sakarya Turkey. Email: gnguler@gmail.com

² Department of Mathematics and Science Education, Ziya Gökalp Faculty of Education, Dicle University, Diyarbakır Turkey. Email: mustafailhan21@gmail.com

³ **Correspondance to:** Ahmet Güneyl (PhD), Social Sciences and Turkish Language Education, Near East University, Nicosia-North Cyprus. Email: ahmet.guneyli@neu.edu.tr

⁴ Department of Measurement and Evaluation in Education, Faculty of Education, Sakarya University, Sakarya Turkey. Email: suleymand@sakarya.edu.tr

Citation: Güler, N., İlhan, M., Güneyl, A., & Demir, S. (2017). An evaluation of the psychometric properties of three different forms of Daly and Miller's writing apprehension test through Rasch analysis. *Educational Sciences: Theory & Practice*, 17, 721–744. <http://dx.doi.org/10.12738/estp.2017.3.0051>

Writing, a basic language skill, is an individual expression of one's knowledge, feelings, thoughts, beliefs, imaginations, and desires in writing (Temizkan, 2014). Writing skills differs from listening, reading, and speaking-which are the other components of language- in a number of ways (Tok & Potur, 2015). Firstly, writing is a skill of description, and in this respect differs from listening and reading (Karatay, 2011). Secondly, although listening and speaking skills are learned to a certain extent in pre-school, writing skills are acquired in formal scholastic education. Formal education's need for developing writing skills makes this skill different from listening and speaking skills (Ungan, 2007). Considering the properties of writing skills, writing skills can be said to be slower and more difficult to develop than other language skills (Güneyli, 2016). Difficulties in developing writing skills and the need for more time to improve it can cause individuals to develop some negative feelings toward it (Yaman, 2010). Upon reviewing the relevant literature (Çocuk, Yanpar Yelken, & Özer, 2016; Daly, 1978; Fathi Huwari & Al-Shboul, 2016; Faigley, Daly, & Witte, 1981; İşeri & Ünal, 2012; Kuşdemir, Şahin, & Bulut, 2016; Yıldız & Ceyhan, 2016), the negative feeling individuals have in relation to writing is mainly found to be writing apprehension.

Writing Apprehension

Writing apprehension, first conceptualized by Daly and Miller (1975; as cited in Smith, 1984), is defined as the anxiety individuals feel in situations where they need to express their feelings and thoughts in writing (Tighe, 1987). Writing apprehension can stem from its complicated nature requiring the use of meta-cognitive skills (Bayat, 2014), from individuals feeling weak and incompetent about their writing, from their negative experiences with writing, or from a lack of reading habits (Zorbaz, 2011). No matter what the source, writing apprehension causes individuals to lose their mental flexibility (Baymur, 1994) and results in failing to generate ideas about what to write (Tiryaki, 2012). Thus, writing apprehension can turn the action of writing into a troublesome and challenging process (Karakaya & Ülper, 2011) that can negatively affect individuals' writing performance (Badrasawi, Zubairi, & Idrus, 2016; Ferguson, 2011; Hassan, 2001).

Measuring Writing Apprehension

Writing apprehension has been a subject of study for a long time in the international literature, and a number of scales exist in the literature developed by various researchers for measuring writing apprehension (Cheng, 2004; Daly & Miller, 1975; Petzel & Wenzel, 1993; Stacks, Boozer, & Lally, 1983). The first scale to measure writing apprehension was the WAT, developed by Daly and Miller (1975). WAT is composed of 26 items, 13 positive and 13 negative. Daly and Miller evaluated the psychometric properties of the test in a sample of university students. In studies performed later

(Richmond & Dickson-Markman, 1985; Singh & Rajalingam, 2012; Zorbaz, 2010), procedures to determine the psychometric properties of the test were repeated with students coming from differing stages of education. As a consequence, the test was found usable for measuring primary-school and high-school students' writing apprehension. Aside from the WAT, another tool for measuring writing apprehension was introduced into the literature by Petzel and Wenzel (1993). The Writing Anxiety Scale (Petzel & Wenzel, 1993) is composed of 103 items placed under nine factors (empathy, expression, evaluation by others, motivation, organization, procrastination, self-esteem, technical skills, and writing anxiety). Another measuring instrument for use in determining writing apprehension is the Second Language Writing Anxiety Inventory (Cheng, 2004). The inventory has three factors (somatic anxiety, cognitive anxiety, and avoidance behavior) and contains 27 items. As is clear from the above-mentioned studies, writing apprehension is an issue that has been studied in the international literature for some time and is still popular in language teaching.

Although studies available in the international literature date back to the 1970s, studies in national literature are as recent as 2010. The study by Zorbaz (2010), which analyses the correlations between writing skills, writing apprehension, and timidity in writing is the first study conducted in Turkey in relation to writing anxiety. In the study, which was conducted with students in their second stage of primary education, Zorbaz (2010) adapted the WAT into Turkish. Hence, the WAT is one of the first measuring tools in this respect in Turkish literature, as it was in the international literature. Other measurement tools for measuring writing apprehension in the Turkish literature include the writing apprehension scale developed by Karakaya and Ülper (2011) for use with prospective teachers and the one developed by Yaman (2010) for use with primary-school students. Although both are single-factor scales, Karakaya and Ülper's (2011) scale contains 35 items, and Yaman's (2010) contains 19.

As evident from the above-mentioned studies, diverse scales are available in both the international and national literature for use in determining individuals' writing apprehension. However, the WAT is the most frequently preferred tool of measurement because it is the first scale developed in the field, can be answered in a short time as it has few items, and can be used with students at all stages of education from primary school to higher education. Despite its widespread use, inconsistency between studies exists concerning the construct validity of the test (Bline, Lowe, Meixner, Nouri, & Pearce, 2001). Daly and Miller (1975), for instance, found that the test had a single factor. Burgoon and Hale (1983), and Shaver (1990), however, found that the test had three factors: ease in writing, enjoying writing, and being rewarded as a consequence of writing. Penley, Alexander, Jernigan, and Henwood (1991), on the other hand, found that it had two factors: difficulty in writing and dislike of writing. Similar to Penley et al. (1991), Bline et al. (2001) found that the

test contained two factors. [Zorbaz \(2010\)](#) found that five out of 26 items did not have adequate factor loadings, and that after these items were removed, the remaining 21 items had four factors: appreciation, prejudice, evaluating apprehension, and sharing what one has written. Furthermore, [Zorbaz \(2010\)](#) stated that WAT can provide a total score regarding students' writing anxiety.

The fact that differing results were obtained in studies using the WAT in relation to the psychometric properties of the test results in the question of what the test's most appropriate structure is to be left unanswered. This situation makes one think that new studies are required for analyzing the psychometric properties of the test. Additionally, the contributions to the literature made by prospective studies analyzing the test's validity and reliability depend on the use of more comprehensive and stronger statistical techniques than those in previous studies. One of these techniques is the Rasch model, which has been frequently used in recent scale-development studies ([Ambiel, Noronha, & Carvalho, 2015](#); [Behizadeh & Engelhard, 2014](#); [Enterline, Cochran-Smith, Ludlow, & Mitescu, 2008](#); [Ricketts, Engelhard, & Chang, 2015](#); [Walker, Engelhard, & Thompson, 2012](#)) and is said to be a powerful method psychometrically ([Brinthaupt & Kang, 2014](#)).

The Rasch Model

Methods based on classical test theory (CTT) are often used in preparing and assessing measurement tools ([Hambleton, Swaminathan, & Rogers, 1991](#)) as they are more practical for mathematical calculations ([Haiyang, 2010](#)), one can work on them with small samples ([You, 2016](#)), and their required assumptions can easily be met ([Hambleton & Jones, 1993](#)). However, CTT-based methods have certain restrictions in the scale development process and the statistical procedures applied to measurements. The restrictions of CTT led researchers to search for methods to help overcome these restrictions. One theory put forward as a result of such investigations is item response theory, also known as latent traits theory. Item response theory represents the family of models that have variations (one-, two-, and three-parameter models; [DeVellis, 2003](#)). The first member of this family is the Rasch model, developed by Danish mathematician George Rasch and also known as a one-parameter model ([Baker, 2001](#)).

The restrictions posed by CTT in determining the validity and reliability of Likert-type scales, and the advantages the Rasch model offers in overcoming these restrictions are as follows. Firstly, the estimations for item parameters in CTT are dependent on the groups to which measurement tools are applied, whereas the ability estimations for participants are dependent on the item sample in the measurement tool. The Rasch model, on the other hand, enables ability calculations independent of item parameters, as well as item parameter estimations independent of ability levels ([Magno, 2009](#)). Secondly, Likert-type scale data are at an ordinal level because the

distance between categories in such scales is not constant throughout the scale; the distance between the first and second categories may differ from the distance between the second and third categories (Harwell & Gatti, 2001). For this reason, Likert-type scales are inappropriate for mathematical operations such as addition, subtraction, multiplication, and division. Yet this situation has not been taken into consideration in CTT, and the scores received from the scale are added up and analyzed through parametric tests. Treating the data in ordinal scales as if they had even intervals can cause scale scores calculated in CTT, as well as the parametric tests applied to those scores, to yield false results (Jamieson, 2004). In contrast to CTT, the original data in the ordinal scale are brought to an interval level in the Rasch model; thusly the restriction can be overcome (Barker, Donovan, Schubert, & Walker, 2016; Tennant, McKenna, & Hagell, 2004). Thirdly, the difficulty levels of all items of a scale (the probability of agreement or disagreement for an item) are regarded as close to each other in CTT ability estimations. The assumption can be misleading that if the differences between difficulty levels of the items represent a different aspect of the targeted quality to be measured, then all items would equally contribute to the scores received from the scale (Anshel, Weatherby, Kang, & Watson, 2009). On the contrary, the Rasch model is based on the assumption that certain items in a scale can be more difficult and that those items can require individuals to have more latent properties that are measured (Curtin, Browne, Staines, & Perry, 2016).

One point where the Rasch model is superior to CTT is related with the conceptualizing measurement errors (You, 2016). Only one standard error value is calculated for all participants in CTT (Zanon, Hutz, Yoo, & Hambleton, 2016), and only one source of error interfering in measurements is considered at a time. Therefore, CTT is unable to provide information about the effects of multiple sources of error on measurement results (Haiyang, 2010). In the Rasch model, however, more than one source of error is considered at a time, the interactions between these sources of error can be determined, and the standard errors of measurements differ at different levels of ability estimation (Embretson & Reise, 2000). Rasch model's ability to actively determine how to rate functions is another advantage. The number of response categories in Likert-type scales is usually determined based on researchers' prior knowledge or on a literature review (You, 2016). While statistically testing whether or not the number of categories in a scale is appropriate for respondents is impossible in analyses performed according to CTT, determining how well scale categories work by examining the tables for category statistics that are among the analysis outputs in the Rasch model is possible (Heesch, Masse, & Dunn, 2006). Finally in Rasch analysis, individuals responding to items other than scale items are also treated as a source of variability that can cause errors in measurement results. In this way, the findings reported in relation to the reliability of measurements are not limited to the items in Rasch analysis; a reliability coefficient is also found for individuals. On the other hand, variability between individuals in CTT is considered to stem from

personal differences, and no information is provided on how to reliably distinguish the individuals responding to scale items (Taşdelen Teker, Güler, & Kaya Uyanık, 2015). Rasch model's superiority over CTT is that it provides information on how to reliably distinguish individuals responding to scale items.

The Purpose and Significance of the Study

Each variable analyzed in a scientific study should be measured correctly and precisely in order to be able to reach valid and reliable conclusions. If any doubts exist as to the measurement of a variable, the results obtained from the measurement, as well as the interpretations of results, will also be doubtful (Tezbaşaran, 1997). In this sense, removing uncertainty in relation to the validity and reliability of scales used in scientific studies is important for raising their scientific quality. Based on such thinking, this study aims to evaluate the psychometric properties of three different versions of the WAT through Rasch analysis. The study will compare the results for the 26-item, single-factor structure of the test's original version with the results of Turkish versions, a 21-item, four-factor structure and a 21-item, single-factor structure. Thus the aim is to determine which of the three listed test versions is the most appropriate for measuring writing apprehension. Considering that the affective dimension of writing education in Turkey is a newly developed field, a comprehensive analysis of the psychometric properties of the WAT, which is one of the most frequently used measurement tools in this field, is predicted to contribute considerably to Turkey's literature. That CTT-based techniques were used in all previous studies analyzing WAT's psychometric properties (Bline et al., 2001; Burgoon & Hale, 1983; Daly & Miller, 1975; Penley et al., 1991; Riffe & Stacks, 1992; Shaver, 1990; Zorbaz, 2010) differentiates the current study from them and turns it into an original study that can also contribute to the international literature.

Method

This part of the study presents information on the study group, data collection tool, and data collection and analysis process.

Study Group

The research for the study was conducted in middle schools (6th-8th graders) in Nicosia, a province in Northern Cyprus. According to data from 2012, 2,854 secondary students attended state schools in Nicosia. The number of students attending private schools, on the other hand, was not available to the researchers. The study group included 720 students chosen at random from state and private schools. Of the participants, 344 (47.78%) were girls and 376 (52.22%) were boys; their ages ranged from 11 – 17 ($M = 12.88$, $SD = 1.10$). Also, 512 (71.11%) participants attended state schools (seven different schools) while 208 (28.89%) attended private schools (three different private

schools). Although 720 students were in the study group, upon checking the data set, many of the measurement tools were found to have been completed incorrectly or incompletely. Measurement tools with unanswered items, with items where more than one alternative had been chosen, or where positive and negative items showed contradictory answers were removed from the data set. Thus, data for 604 students remained in the data set, and the analyses were performed on these.

Data Collection Tool

The WAT, adapted into Turkish by Zorbaz (2010), was used as the data collection tool. The original and Turkish versions' characteristics of the test are as follows:

The Original Form of WAT. A tool of 63 items was first formed. Then, the 63-item draft of the measurement tool was administered to 164 university students in a five-point Likert-type scale, and factor analysis was performed on the data. Through the varimax rotation and factor analysis performed on the principal components' technique, the test items were found to be divided into two factors. On examining the items' distribution according to factors, a distinction was found between positive and negative items. Namely, one of the factors contained only positive items whereas the other factor contained only negative items. However, because this structure was not found to be interpretable, factor analysis was fixed to a single dimension and analysis was repeated. Following analysis, items with factor loadings below 0.57 were removed from the test, and the analysis was repeated again. In this way, a single factor structure is reached with 26 items (13 positive and 13 negative) whose factor loadings are greater than 0.60 and 46% of the total variance is explained. Split-half and test-retest methods were used in order to evaluate the test's reliability. The split-half reliability coefficient was found to be 0.94. The correlation between two applications performed at a one-week interval for test-retest reliability was found as 0.92.

Turkish Version of WAT (Zorbaz, 2010). The transactions with the Turkish version of WAT were performed with 450 primary-school second-level students in total, 234 (52%) of whom were girls and 216 (48%) of whom were boys. The test was translated into Turkish by five experts with a good command of the English language. After the translation, a linguistic equivalency study was conducted with 36 students from the English Language Teaching Department. Having concluded that linguistic equivalence had been attained between the English and Turkish versions, factor analysis was performed to demonstrate the construct validity of the Turkish version of the scale. Following factor analysis, the scale was found to explain 53% of the total variance through four dimensions. The contributions of each dimension to the total explained variance in this four-dimensional structure, the number of items each dimension contains, factor loadings, and reliability are shown in Table 1, which also shows a sample item from each dimension.

Table 1
Information on the Construct Validity and Reliability of the Turkish Version of WAT

Factors	Number of items	Explained variance	Range of Factor loadings	Reliability	Sample items
AP	5	32.18%	0.67-0.79	0.84	I like to write down my ideas.
PR	7	9.14%	0.54-0.72	0.79	I expect to do poorly in composition classes even before I enter them.
EA	6	6.30%	0.50-0.67	0.69	I don't like my compositions to be evaluated.
SW	3	5.41%	0.62-0.79	0.68	I like to have my friends read what I have written.

AP: Appreciation, PR: Prejudice, EA: Evaluating apprehension, SW: Sharing what one has written

As is clear from Table 1, the Turkish version of WAT contains 21 items, which is different from the original version. This situation stems from either the five items in the original version of the test being outside of the four-dimensional structure arising in Turkish culture or their having high factor loads in more than one dimension. Although the Turkish version of WAT has a four-dimensional structure, [Zorbaz \(2010\)](#) demonstrated that the scale could also be used one dimensionally and found Cronbach's alpha internal consistency to be .90 for the whole test.

Data Collection and Analysis

The data were collected from students in the classroom setting. Prior to applying the measurement tool, the students were informed of the purpose of the research and were made aware of the extreme importance of answering the test carefully so that correct conclusions could be reached. They were also informed that participating in the research was not obligatory, thus ensuring that the study group was made up of volunteers. Variables such as gender, age, and grade level were added at the top of the data collection tool; participants' demographic information was also collected. The students took approximately 15-20 minutes to answer the items in the data collection tool. As the 26-item form was initially administered to the students, the data obtained from this application was used to analyze the 26-item, one-dimensional structure; the 21-item, one-dimensional structure; and the 21-item, four-dimensional structure. Thus, collecting data separately for each form of the WAT was not considered necessary.

A preliminary check was performed to remove incomplete (unanswered items) or inconsistent (same answers to positive and negative items) measurement tools from the data set. Consequently, measurement tools for 116 students were excluded from the study. After reverse scoring the negative items, the data were prepared for analyses. The psychometric properties of the data were assessed using Rasch analysis. Each source of variability capable of influencing the measurement results is called a *facet* in the Rasch model, and Rasch analysis can be performed as two-faceted or multi-faceted according to the number of sources of variability. For Likert-type scales, the sources of variability

that can influence measurement results are limited to items and individuals. Therefore, this study uses two-faceted Rasch analysis. The analyses were done according to the rating scale model through the FACETS package program and joint (unconditional) maximum-likelihood estimations were used. Whether or not the assumptions of the Rasch model had been met was tested prior to interpreting the analysis results. The Rasch analysis has three assumptions: unidimensionality, local independence, and model-data fit (DeMars, 2010). However, because these assumptions function parallel with each other, testing each separately is unnecessary. Attaining model-data fit in Rasch analysis indicates that the assumption of unidimensionality has been met (Lee, Peterson, & Dixon, 2010), and meeting the assumption of unidimensionality means that the condition of local independence has been met (Hambleton et al., 1991). Therefore, the basic assumption to test is model-data fit.

Fit statistics are the reference in assessing model-data fit in Rasch analysis. Fit statistics provide information about how well the observed and the expected values for each source of variability overlap. These statistics, which can be divided into infit and outfit statistics, have a “0” standard-error value and a “1” expected value. Fit statistics with a value of 1 indicate that model-data fit is perfect. However, having perfect fit is often impossible under the actual conditions of measurement (Brentari & Golia, 2008). For this reason, the acceptable interval for fit statistics should be determined. Even though various suggestions have been made by different researchers in relation to acceptable intervals for fit statistics, the most commonly accepted criterion is the 0.5 – 1.5 interval (Wright & Linacre, 1994). Accordingly, when the majority of items in a measurement tool are within the acceptable interval of 0.5 and 1.5, the interpretation for the situation is that model-data fit has been achieved (Brinthaupt & Kang, 2014). On revising the results for Rasch analysis, all of the fit statistics were found to remain within acceptable limits (see the section on findings). Therefore, the assumption of model-data fit can be said to be met. The fit between model and data means that the assumptions of unidimensionality and local independence have also been met.

Having found that the assumptions are met, the psychometric properties of the three different WAT models were compared. Fit statistics reported in Rasch analysis outputs, point-biserial correlation coefficients for the items; separation ratios, reliability coefficients, and chi-squared values for the facets of item and person; and category statistics for the rating used in the scale were taken into consideration for this study’s purposes. Fit statistics were first examined for all three models. The decision was made to remove any items with fit statistics below 0.5 or over 1.5 (Anshel et al., 2009).

After fit statistics, point-biserial correlations for the items were examined. These correlations provide information about whether all elements of a facet function in the same way. In this aspect, point-biserial correlation coefficients in the Rasch model

are considered to be counterparts to Pearson's correlations in CTT (Linacre, 2014). There are no defined criteria for point-biserial correlations in the literature on Rasch analysis. Therefore, the researchers themselves decided on the criterion for which to base point-biserial correlations. Taking into consideration that these correlation coefficients are counterparts to Pearson's correlations in CTT, the criterion value was accordingly taken as 0.30 because, as is commonly known, there is a condition in CTT that item total correlation should be greater than 0.30 for item discrimination (Field, 2009).

The chi-square values, reliability coefficients, and separation ratios for the facets of item and person were assessed after fit statistics and point-biserial correlations. Chi-squared demonstrates whether any significant differences exist between the elements of a facet. The significance of the chi-square for the facet of person means that persons responding to the measurement tool with qualities that have been measured at differing levels are significantly distinct. In a similar vein, the significance of the chi-square calculated for the item facet shows that significant differences exist between the items' difficulty levels in the measurement tool. The reliability coefficient and separation ratio provide statistical information on how reliably the facet elements have been distinguished. According to Bond and Fox (2007), the reliability coefficient in Rasch analysis is the counterpart of Cronbach's alpha internal consistency coefficient in CTT. Setting out from this fact, Walker et al. (2012) said that values of .70 can be taken as a lower limit for the reliability coefficient in Rasch analysis, just as for Cronbach's alpha for internal consistency. The separation ratio also presents the same information as the reliability coefficient does but is reported in a different metric (Çetin & İlhan, 2017). While reliability coefficients can take on values between 0 and 1, separation ratios can take on values between 1 and infinity (Sudweeks, Reeveb, & Bradshawc, 2005). Separation values of 2 or higher are considered sufficient for declaring facet elements to be effectively distinguished (Linacre, 2012).

Finally, category statistics were examined in the Rasch analysis outputs. Category statistics provide information on whether the type of rating employed in a scale functions effectively. When the measures shown in the table of category statistics increase monotonically, the outfit statistics are in the 0.5 – 1.5 interval, and there are at least 10 observations for each category of a scale, then the interpretation is that the type of adopted rating works smoothly (Linacre, 2014). The points where measures do not increase in parallel to scale categories are shown with the * symbol. Such a result indicates that scale categories are not distinguished well by respondents and that the response alternatives should be combined.

Findings

This section presents the obtained findings. First, fit statistics for the items in all three forms and point-biserial correlations were analyzed. These obtained findings are shown in Table 2.

Table 2
Fit Statistics and Correlations for Items in the Three Forms of WAT

Form	Item No.	Measure	Infit MnSq	Outfit MnSq	Corr. PtBis	Item No.	Measure	Infit MnSq	Outfit MnSq	Corr. PtBis	Overall Fit Statistics
26-item, one-dimensional	11	-0.30	1.17	1.34	0.31	114	0.09	0.74	0.79	0.41	Infit: 1.00 Outfit: 1.05
	12	-0.08	1.04	1.11	0.30	115	-0.09	0.91	0.90	0.51	
	13	0.17	1.00	1.03	0.41	116	0.07	0.97	0.97	0.46	
	14	-0.23	1.07	1.15	0.35	117	0.03	0.95	0.96	0.47	
	15	-0.31	0.99	0.95	0.44	118	-0.13	1.07	1.10	0.46	
	16	-0.03	1.08	1.24	0.34	119	-0.02	0.78	0.76	0.54	
	17	0.20	1.33	1.4	0.31	120	0.26	1.09	1.11	0.36	
	18	-0.14	0.97	1.12	0.40	121	0.11	0.90	0.90	0.37	
	19	0.25	1.39	1.49	0.27	122	-0.16	0.92	0.90	0.53	
	110	-0.03	0.92	0.89	0.54	123	0.12	0.76	0.76	0.52	
	111	-0.04	0.79	0.82	0.54	124	0.03	1.14	1.24	0.35	
	112	0.37	1.12	1.15	0.35	125	0.09	1.18	1.42	0.32	
	113	-0.20	0.89	0.86	0.48	126	-0.04	0.92	0.93	0.48	
21-item, one-dimensional	11	-0.30	1.17	1.29	0.31	116	0.08	1.00	1.02	0.43	Infit: 1.00 Outfit: 1.04
	12	-0.08	1.04	1.11	0.30	117	0.04	0.93	0.94	0.48	
	13	0.17	0.98	1.05	0.42	118	-0.12	1.10	1.12	0.43	
	14	-0.22	1.09	1.17	0.33	119	-0.01	0.77	0.74	0.54	
	15	-0.31	1.01	0.99	0.41	120	0.27	1.07	1.08	0.37	
	17	0.26	1.37	1.45	0.29	122	-0.16	0.96	0.94	0.50	
	110	-0.03	0.91	0.88	0.54	123	0.13	0.78	0.79	0.49	
	112	0.38	1.11	1.14	0.36	124	0.03	1.15	1.23	0.34	
	113	-0.20	0.89	0.89	0.47	125	0.10	1.20	1.41	0.30	
	114	0.10	0.73	0.77	0.42	126	-0.03	0.95	0.96	0.46	
	115	-0.09	0.88	0.87	0.53						
26-item, four-dimensional	AP-13	0.28	1.17	1.22	0.54	PR-17	0.24	1.31	1.31	0.38	Infit; AP: 1.00 EA: 1.00 PR: 1.00 SW: 0.99 Outfit; AP: 1.01 EA: 1.03 PR: 1.03 SW: 0.98
	AP-110	-0.08	0.92	0.90	0.66	PR-116	0.07	0.99	1.00	0.48	
	AP-115	-0.19	0.96	0.98	0.63	PR-118	-0.18	0.90	0.89	0.58	
	AP-117	0.04	1.06	1.07	0.54	PR-122	-0.22	0.83	0.79	0.59	
	AP-119	-0.05	0.90	0.89	0.62	PR-123	0.13	0.94	1.04	0.44	
	EA-11	-0.15	1.14	1.21	0.26	PR-124	0.02	1.12	1.20	0.42	
	EA-12	0.10	1.09	1.13	0.21	PR-126	-0.06	0.93	0.96	0.50	
	EA-14	-0.06	0.90	0.91	0.40	SW-112	0.23	0.97	0.94	0.41	
	EA-15	-0.16	0.91	0.91	0.40	SW-114	-0.26	1.02	1.06	0.41	
	EA-113	-0.03	0.79	0.81	0.46	SW-120	0.03	0.97	0.95	0.30	
	EA-125	0.31	1.17	1.18	0.24						

AP: Appreciation, PR: Prejudice, EA: Evaluating apprehension, SW: Sharing what one has written (Items 6, 8, 9, 11, and 21 are in the 26 item-single factor form of the scale but they are not in the 21 item-single factor and 21 item-four factor structures of the test.)

The findings shown in Table 2 demonstrate that the infit and outfit mean squares are within acceptable limits ($0.5 < MnSq < 1.5$; Wright & Linacre, 1994) in all three forms of WAT. Accordingly, one can say no item needs to be removed from any of the tests, and that all the items in the tests serve to measure writing apprehension. According to Table 2, all correlations were found to be above 0.30 in the 21-item, one-dimensional structure. Point-biserial correlation coefficients were also found to be above the 0.30 criterion for items, except for Item 9 in the 26-item, one-dimensional form. The correlation coefficient calculated for Item 9 was .27, which is less than .30. However, because the reported value is quite close to the criterion value, Item 9 can also be said to function the same as the other items in the scale. An examination of the correlations in the 21-item, four-dimensional form of WAT shows that correlation coefficients calculated for all items in terms of apprehension, prejudice, and sharing what one has written are greater than .30. In terms of evaluating apprehension, however, point-biserial correlation coefficients for three of six items were found not to meet the 0.30 criterion. Thus, one can say items in the factor of evaluating apprehension don't function the same. Upon examining the values shown in Table 3 for reliability, separation ratios, and chi-squares calculated for the person facet, the lowest values are clearly in the factor of evaluating apprehension in the 21-item, four-dimensional form of WAT. In addition to fit statistics and correlations for the items, Table 3 also shows the reliability, separation ratio, and chi-square values calculated for the facets of item and person in three different structures.

Table 3
Reliability, Separation Ratio and Chi Square Values for the Facets of Item and Person in Three Forms of WAT

Form	Item Facet			Person Facet			
	Reliability	Separation Ratio	Chi-square	Reliability	Separation Ratio	Chi-square	
26-item, one-dimensional	.95	4.56	$x^2 = 538.0^*$.87	2.64	$x^2 = 3567.10^*$	
21-item, one-dimensional	.96	4.87	$x^2 = 489.7^*$.85	2.42	$x^2 = 3013.40^*$	
21-item, four-dimensional	AP	.92	3.44	$x^2 = 52.10^*$.79	1.92	$x^2 = 2073.30^*$
	EA	.95	4.21	$x^2 = 97.10^*$.64	1.33	$x^2 = 1171.20^*$
	PR	.94	3.89	$x^2 = 96.50^*$.76	1.79	$x^2 = 1754.20^*$
	SW	.96	5.04	$x^2 = 52.5^*$.66	1.39	$x^2 = 1184.40^*$

* $p < .01$; AP: Appreciation, PR: Prejudice, EA: Evaluating apprehension, SW: Sharing what one has written

According to Table 3, the reliability, separation ratio, and chi-square values for the facets of item and person are very close in the 26-item, one-dimensional form and the 21-item, one-dimensional form. The calculated chi-square values exhibit significant differences between item difficulty levels in both models and that students with differing writing apprehension can be significantly distinguished. Upon examining

the reliability coefficient and separation ratio, items as well as students become clearly distinguishable with high reliability, whether the scale is the 26-item, one-dimensional or the 21-item, one-dimensional test.

On examining the findings for the 21-item, four-dimensional structure, the results for the facet of item were found to be similar to the 26-item, one-dimensional and the 21-item, one-dimensional forms in terms of item-difficulty levels. Significant differences were also found in terms of item difficulty levels for the 21-item, four-dimensional form as well as the other two forms; the items were distinguished with high reliability. However, considerable differences exist among the results obtained for the 21-item, four-dimensional form, the 26-item, one-dimensional form, and the 21-item, one-dimensional form in terms of the person facet. Even though students with different writing apprehensions were distinguished significantly in the 21-item, four-dimensional form, the reliability of estimations on their levels of writing apprehension were found to be low. In contrast with the other two tested forms, the 21-item, four-dimensional structure's reported separation ratios do not meet the criterion value of 2 (Linacre, 2012) for considering these statistics in any factor. In a similar vein, the reliability values calculated for the facet of person were also found to be below the .70 criterion (Bond & Fox, 2007) in the factors of evaluating apprehension and sharing what one has written.

The reliability, separation ratio, and chi-square values shown in Table 3 make it clear that the 21-item, four-dimensional form is not appropriate for use in measuring students' writing apprehension and that students' apprehension can be measured more reliably through the 26-item, one-dimensional form or the 21-item, one-dimensional form. However, the findings shown in Table 3 are insufficient on their own for deciding which of the two forms is more appropriate for measuring students' writing apprehension. Considering the fact that the 21-item, one-dimensional form, despite containing fewer items, yields measurement results similar to the 26-item, one-dimensional form psychometrically, the 21-item form is considered to be a more preferable model. But before deciding on which of the two models is more suitable, the fact that ability estimations reported in relation to students' writing apprehension in both models are similar needs to be reported. Whether any differences exist between ability estimations made in both models was analyzed through the paired samples *t*-test, and these findings are shown in Table 4.

Table 4
Paired Samples t-test Results for Comparing Estimations Reported for the 26-Item, One-Dimensional and the 21-Item, One-Dimensional Forms of WAT

Form	<i>N</i>	<i>M</i>	<i>SD</i>	<i>df</i>	<i>r</i>	<i>t</i>
26-item, one-dimensional	604	3.18	0.57	603	.97**	1.36*
21-item, one-dimensional		3.26	0.62			

***p* < .01, **p* > .05.

Correlations shown in Table 4 indicate a high relative agreement between the estimations made from the two different WAT forms [$r = .97, p < .01$]. According to Table 4, no significant differences exist between the estimations reported in the 26-item, one-dimensional and the 21-item, one-dimensional forms of WAT [$t_{(603)} = 1.36, p > .05$]. The statistical insignificance of the paired samples *t*-test results indicates absolute agreement between estimations reported in both forms of the scale. Accordingly, no matter which of the two tests is used, one can say that there will be no difference in the decisions to be made in relation to students' writing apprehension. Therefore, saying that the 21-item, one-dimensional form that contains a smaller number of items is more useful and a better choice would not be wrong.

Although the 21-item, one-dimensional form of WAT is more appropriate than the other two forms, that this form is inadequate in some ways at discriminating between students with different writing apprehension should not be ignore. That is to say, the range values for item-difficulty levels [$(0.38) - (-0.31) = 0.69$] are lower than the range values for students' ability levels on the 21-item, one-dimensional form of the scale. Therefore, items in the 21-item, one-dimensional form that are suitable for students with medium levels of writing apprehension can be said, but no items exist that are suitable for students with low or high levels of writing apprehension. This is also evident in the variable map shown in Figure 1. According to the variable map, while students distribute extensively in terms of their writing apprehension, items have narrower distribution in terms of difficulty levels. In other words, no items in the 21-item form of WAT correspond to the apprehension levels of students located at the lower and upper ends of the variable map. Thus, the number of errors merged into the estimations for students with low or high levels of writing apprehension is greater than those with medium levels of writing apprehension. For example, the standard error calculated for a student with a writing apprehension of 4.94 is 1.82, whereas the standard error calculated for a student with a writing apprehension of 0.01 is 0.18. Therefore, when items whose difficulty levels are higher or lower than the existing items of the test are added to WAT, one can say that more accurate estimations for students' writing apprehension will occur (especially for students with high or low levels of writing apprehension).

Recommending adding items to the 21-item form while not preferring the 26-item form may seem contradictory. Clarifying this is the fact that the range calculated for the difficulty levels of items in the 26-item, one-dimensional form [$(0.37) - (-0.30) = 0.67$] approach the range calculated for the difficulty levels of items in the 21-item, one-dimensional form (0.69). Put more clearly, the extra five items in the 26-item, one-dimensional form of the scale are no different from the other items of the scale in terms of difficulty level. Hence, the items included in the 26-item form not in the 21-item, one-dimensional form do not contribute to more accurate estimations for

students with high or low levels of writing apprehension. For this reason, the 21-item, one-dimensional form of the test is a better choice and should be the base for adding more difficult and easier items to the scale.

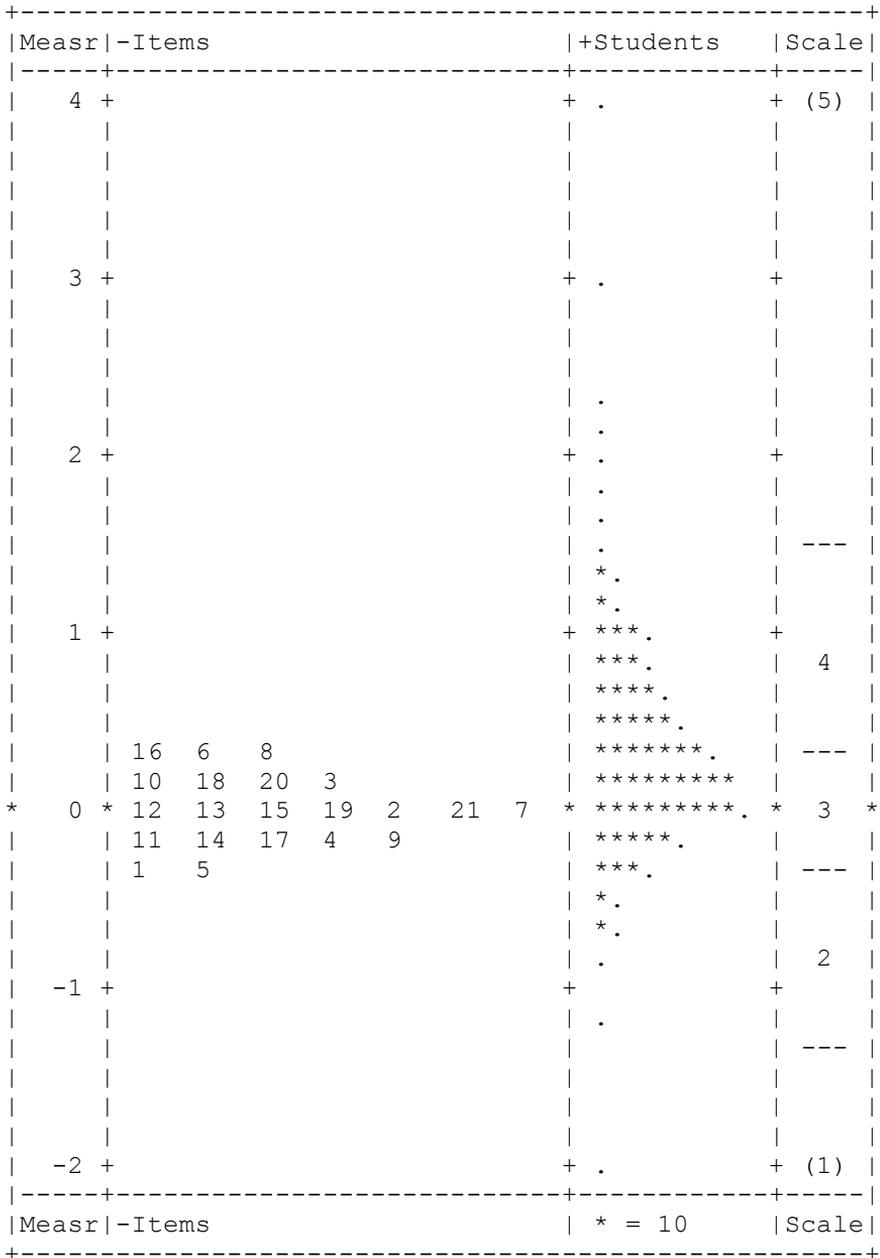


Figure 1. Variable map for the 21-item, one-dimensional form of WAT.

This study analyzed category statistics for each model in addition to fit statistics and point biserial correlation coefficients for the items, and reliability, separation ratio and Chi square values for the facets of item and person. In this way, the study aimed to determine whether or not the test categories could be discriminated by responders in the five-pointed rating [*Strongly disagree* (1) → *Strongly agree* (5)] adopted. Table 5 shows the category statistics for the 26-item, one-dimensional, 21-item, one dimensional, and 21-item, four-dimensional forms of WAT.

Table 5
Category Statistics for the Three Different Forms of WAT

Form	Category	Frequency	Percent	Average Measure	Expected Measure	Outfit MnSq	
26-item, one dimension	1	1,574	10%	-0.12	-0.17	1.2	
	2	2,329	15%	0.02	0.02	1.0	
	3	3,770	24%	0.17	0.20	0.8	
	4	4,219	27%	0.38	0.41	0.9	
	5	3,786	24%	0.72	0.69	1.0	
21-item, one dimension	1	1,251	10%	-0.11	-0.16	1.2	
	2	1,876	15%	-0.01	0.01	1.0	
	3	3,011	24%	0.17	0.20	0.8	
	4	3,414	27%	0.40	0.42	0.9	
	5	3,090	24%	0.77	0.73	1.0	
21-item, four-dimension	AP	1	241	9%	-.93	-1.05	1.3
		2	443	16%	-.51	-0.44	1.0
		3	706	25%	0.08	0.14	0.7
		4	864	31%	0.87	0.81	0.8
		5	536	19%	1.66	1.68	1.1
	EA	1	245	7%	-0.08	-0.20	1.3
		2	474	13%	0.01	0.05	0.9
		3	788	22%	0.28	0.31	0.9
		4	1,036	29%	0.62	0.62	0.9
		5	985	28%	1.06	1.04	1.0
PR	1	441	11%	-0.52	-0.58	1.2	
	2	672	16%	-0.17	-0.17	1.0	
	3	952	23%	0.11	0.16	0.8	
	4	1,087	26%	0.55	0.55	0.9	
	5	985	24%	1.17	1.15	1.1	
SW	1	208	12%	-1.07	-1.11	1.1	
	2	302	18%	-0.72	-0.59	0.8	
	3	535	31%	-0.02	-0.05	0.8	
	4	451	26%	0.69	0.57	0.9	
	5	211	12%	1.16	1.33	1.2	

According to Table 5, average measurements increase from the lower end (*strongly disagree*) to the upper end (*strongly agree*) of the five-point rating. According to the table, the distribution of frequencies into scale categories is regular, and outfit mean squares are in the acceptable 0.5 – 1.5 interval (Wright & Linacre, 1994). Accordingly, all the assumptions have been met for saying that the adopted rating in

the test functions smoothly. In other words, the five-point rating system used in the test can be said to work effectively in all three forms of WAT.

Discussion, Conclusions, and Recommendations

This study evaluates the psychometric properties of three different forms of WAT through Rasch analysis. Accordingly, the fit statistics; correlation coefficients; and the reliability, separation ratio, and chi square values for the facets of item and person calculated for the 26-item, one-dimensional; 21-item, one-dimensional; and 21-item, four-dimensional forms of WAT were compared. The findings demonstrate that fit statistics are within the acceptable 0.5 – 1.5 interval for all three WAT forms. Fit statistics calculated as less than 0.5 for an item indicates that the item does not provide information distinct from other items, this also indicates that the assumption of local independence may have been violated (Chan, Chien, Su, & Lin, 2009). Fit statistics greater than 1.5, on the other hand, mean that these items with higher fit statistics do not measure the same structure as other items in the scale do (Engelhard, 2011). Accordingly, one can state that the item scales serve to measure students' writing apprehension in all three forms of the test and that each item can be answered independent of the others.

On examining the point-biserial correlations in the Rasch analysis outputs, no problems were found concerning item correlations in the 26-item, one-dimensional and the 21-item, one-dimensional forms. Based on this finding, all items can be said to function the same way in the one-dimensional forms of WAT. On checking the correlations in the 21-item, four-dimensional form of the test, the correlation coefficients for three of the six items in the dimension of evaluating apprehension were found to be below the 0.30 criterion. Thus, all the items in the dimension of evaluating apprehension can be said to not function the same. This finding on point-biserial correlation coefficients overlaps with the fit statistics calculated for the items because the items with correlations greater or less than 0.30 in the sub-scale of evaluating apprehension form two different sets in terms of fit statistics. While fit statistics calculated for items with high correlations range between 0.79 and 0.91 (less than 1.00, which shows perfect fit), fit statistics reported for items with correlations below 0.30 are between 1.09 and 1.21 (greater than 1.00, which shows perfect fit).

Significant differences were found between items in terms of difficulty levels on all three WAT forms according to the chi-square values calculated for the facets of item and person in Rasch analysis and the students were significantly distinguished. On examining the reliability and separation ratio values reported for the item facet, reliability coefficients are found to be above .70 and the separation ratio above the criterion of 2 in the 26-item, one-dimensional; 21-item, one-dimensional; and 21-

item, four-dimensional forms. Thus, item reliability was concluded to be sufficient in all three forms of the scale. The reliability coefficients and separation ratios for the facet of person show that students were distinguished with high reliability in terms of writing apprehension on the 26-item, one-dimensional and the 21-item, one-dimensional forms. On the other hand, the reliability coefficient and separation ratio were found to be lower on the 21-item, four-dimensional form. The separation ratio was found to be lower on all sub-scales, and reliability coefficients were found to be lower than the criterion values in the sub-scales of evaluating apprehension and sharing what one has written (2 and 0.70, respectively) in the 21-item, four-dimensional form. Based on this finding, the rate of errors merging into estimations for students' writing apprehension can be said to increase when using the 21-item, four-dimensional form. The fewer number of items in the sub-scales in the four-dimensional form is thought to be a possible factor leading to this result.

Point-biserial correlation coefficients, reliability coefficients, and separation ratios calculated for the facet of person show that the 21-item, four-dimensional form of the scale would not be a good choice and that the one-dimensional forms with 21 and 26 items were similar in terms of psychometric properties. Another quality apart from validity and reliability is known to be the usefulness of measurement tools (Güler, 2012). Therefore, the 21-item, one-dimensional form of WAT is more preferable than the 26-item, one-dimensional form. The statistical insignificance of the result for the paired samples *t*-test performed with ability estimations on the 26-item, one-dimensional and 21-item, one-dimensional forms of the scale shows that the 21-item, one-dimensional model can produce estimations parallel to those in the 26-item, one-dimensional model with fewer items. This finding is similar to that obtained by Zorbaz (2010) when adapting WAT into Turkish. The five items available in the original WAT form were removed from the scale by Zorbaz (2010) because they did not apply to Turkish culture. Removing these five items from the scale does not cause any loss in information about students' writing apprehension, which was confirmed by the findings obtained in this study through a different sample in Cyprus, another area where Turkish is spoken as the native language.

Although the most appropriate form of the scale is the one with 21 items and one dimension according to the results obtained in this study, some aspects of this form need developing. In item response theory, the most accurate estimations can be made at times when individuals' ability levels match the items' difficulty levels (Crocker & Algina, 1986). Therefore, easy items (high probability of students' agreement) can yield more accurate estimations for participants with low ability levels, whereas difficult items (low probability of student agreement) can yield more accurate estimations for participants with high ability levels (Baker, 2001). The absence of items in WAT corresponding to students' with very low or very high apprehension

leads to an increase in the amount of errors. This result means that items with difficulty levels (lower or higher probability of student agreement) other than what is on the test should be added to the scale.

Another finding obtained in this study is how functional the five-point rating used in WAT was. Following Rasch analysis, the five-point rating was found to work effectively on all three forms of WAT. This finding parallels theoretical knowledge and the findings obtained in studies available in the literature. For instance, in a book where the author describes the scale development process, Tezbaşaran (1997) states that three-, five-, or seven-point scoring can be used in Likert-type scales, but that the most appropriate number is five (as also indicated by Likert, 1932). Similarly, İlhan and Güler (2016) analyzed the effects of the number of response categories on validity and reliability by using Rasch analysis and found that five-point scoring is more appropriate than three- or seven-point scorings for Turkish culture.

In conclusion, the 21-item, one-dimensional form was found to be the most appropriate model for WAT, and five-point scoring was found to work the most effectively. However, considering the fact that the psychometric properties of measurements and the most appropriate number of categories can vary according to such demographic properties as cultural properties, differences between native language and foreign language, age, and level of education, one may suggest that a similar study should be conducted in different cultures, at different stages of education, and with different age groups.

References

- Ambiel, R. A. M., Noronha, A. P. P., & Carvalho, L. F. (2015). Analysis of the Professional Choice Self-efficacy Scale using the Rasch-Andrich Rating Scale Model. *International Journal for Educational and Vocational Guidance, 15*(3), 205–219. <http://dxdoi.org/10.1186/s41155-016-0021-0>
- Anshel, M. H., Weatherby, N. L., Kang, M., & Watson, T. (2009). Rasch calibration of a unidimensional perfectionism inventory for sport. *Psychology of Sport and Exercise, 10*(1), 210–216. <http://dxdoi.org/10.1016/j.psychsport.2008.07.006>
- Badrasawi, K. J. I., Zubairi, A., & Idrus, F. (2016). Exploring the relationship between writing apprehension and writing performance: A qualitative study. *International Education Studies, 9*(8), 134–143. <http://dxdoi.org/10.5539/ies.v9n8p134>
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
- Barker, B. A., Donovan, N. J., Schubert, A. D., & Walker, E. A. (2016). Using Rasch analysis to examine the item-level psychometrics of the Infant-Toddler Meaningful Auditory Integration scales. *Speech, Language and Hearing, 10*. <http://dxdoi.org/10.1080/2050571X.2016.1243747>
- Bayat, N. (2014). The effect of the process writing approach on writing success and anxiety. *Educational Sciences: Theory & Practice, 14*, 1123–1141.

- Baymur, F. (1994). *Genel psikoloji* [General psychology]. İstanbul, Turkey: İnkılap Bookstore.
- Behizadeh, N., & Engelhard, G. (2014). Development and validation of a scale to measure perceived authenticity in writing. *Assessing Writing, 21*, 18–36. <http://dxdoi.org/10.1016/j.asw.2014.02.001>
- Bline, D., Lowe, D. R., Meixner, W. F., Nouri, H., & Pearce, K. (2001). A research note on the dimensionality of Daly and Miller's writing apprehension scale. *Written Communication, 18*(1), 61–79. <http://dxdoi.org/10.1177/0741088301018001003>
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.
- Brentari, E., & Golia, S. (2008). Measuring job satisfaction in the social services sector with the Rasch model. *Journal of Applied Measurement, 9*(1), 45–56.
- Brinthaup, T. M., & Kang, M. (2014). Many-faceted Rasch calibration: An example using the Self-talk Scale. *Assessment, 21*(2) 241–249. <http://dxdoi.org/10.1177/1073191112446653>
- Burgoon, J. K., & Hale, J. L. (1983). Research note on the dimensions of communication reticence. *Communication Quarterly, 31*, 238–248. <http://dxdoi.org/10.1080/01463378309369510>
- Çetin, B., & İlhan, M. (2017). An analysis of rater severity and leniency in open-ended mathematic questions rated through standard rubrics and rubrics based on the SOLO taxonomy. *Education and Science, 42*(189), 217–247. <http://dxdoi.org/10.15390/EB.2017.5082>
- Chan, A. L. F., Chien, T. W., Su., C. Y., & Lin, S. J. (2009). Validation of a shortened Taiwanese version of an asthma quality-of-life questionnaire by Rasch model. *Allergy and Asthma Proceedings, 30*(2), 171–180. <http://dxdoi.org/10.2500/aap.2009.30.3200>
- Cheng, Y. S. (2004). A measure of second language writing anxiety: Scale development and preliminary validation. *Journal of Second Language Writing, 13*(4), 313–335. <http://dxdoi.org/10.1016/j.jslw.2004.07.001>
- Çocuk, H. E., Yanpar Yelken, T., & Özer, Ö. (2016). The relationship between writing anxiety and writing disposition among secondary school students. *Eurasian Journal of Educational Research, 63*, 335–352. <http://dxdoi.org/10.14689/ejer.2016.63.19>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart & Winston, Inc.
- Curtin, M., Browne, J., Staines, A., & Perry, I. J. (2016). The early development instrument: An evaluation of its five domains using Rasch analysis. *BMC Pediatrics, 16*, 10. <http://dxdoi.org/10.1186/s12887-016-0543-8>
- Daly, J. A. (1978). Writing apprehension and writing competency. *The Journal of Educational Research, 72*(1), 10–14. Retrieved from <http://www.jstor.org/stable/27537168>
- Daly, J. A., & Miller, M. D. (1975). Apprehension of writing as a predictor of message intensity. *The Journal of Psychology: Interdisciplinary and Applied, 89*(2), 175–177. <http://dxdoi.org/10.1080/00223980.1975.9915748>
- DeMars, C. (2010). *Item response theory*. Oxford, UK: Oxford University Press.
- DeVellis, R. F. (2003). *Scale development: Theory and applications*. Newbury Park, CA: Sage.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum & Associates, Inc.

- Engelhard, G. (2011). Evaluating the bookmark judgments of standard-setting panelists. *Educational and Psychological Measurement, 71*(6), 909–924. <http://dxdoi.org/10.1177/0013164410395934>
- Enterline, S., Cochran-Smith, M., Ludlow, L. H., & Mitescu, E. (2008). Learning to teach for social justice: Measuring change in the beliefs of teacher candidates. *The New Educator, 4*(4), 267–290. <http://dxdoi.org/10.1080/15476880802430361>
- Faigley, L., Daly, J. A., & Witte, S. P. (1981). The role of writing apprehension in writing performance and competence. *Journal of Educational Research, 75*(1), 16–21. <http://dxdoi.org/10.1080/00220671.1981.10885348>
- Fathi Huwari, I., & Al-Shboul, Y. (2016). Student's strategies to reduce writing apprehension (A case study on Zarqa University). *Mediterranean Journal of Social Sciences, 7*(3), 283–290. <http://dxdoi.org/10.5901/mjss.2016.v7n3s1p283>
- Fergusson, A. J. (2011). *Writing anxiety and writing performance: A study of Barbadian students* (Unpublished doctoral dissertation). The University of West Indies, Mona, Kingston, Jamaica. Retrieved from http://uwispace.sta.uwi.edu/dspace/bitstream/handle/2139/13415/AnnJacquelineFergusson_AB.pdf?sequence=1
- Field, A. (2009). *Discovering statistics using SPSS*. London, UK: Sage.
- Güler, N. (2012). *Eğitimde ölçme ve değerlendirme* [Measurement and evaluation in education]. Ankara, Turkey: Pegem Academy Publishing.
- Güneylü, A. (2016). Analyzing writing anxiety level of Turkish Cypriot students. *Education and Science, 41*(183), 163–180. <http://dxdoi.org/10.15390/EB.2016.4503>
- Haiyang, S. (2010). An application of classical test theory and many facet Rasch measurement in analyzing the reliability of an English test for non-English major graduates. *Chinese Journal of Applied Linguistics, 33*(2), 87–102. Retrieved from <http://www.celea.org.cn/teic/90/10060807.pdf>
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 38–47. <http://dx.doi.org/10.1111/j.1745-3992.1993.tb00543.x>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research, 71*(1), 105–131. <http://dxdoi.org/10.3102/00346543071001105>
- Hassan, B. A. (2001). *The relationship of writing apprehension and self-esteem to the writing quality and quantity of EFL university students* (ERIC Document [ED459971]). Washington, DC: ERIC Institute of Education Sciences. Retrieved from <https://eric.ed.gov/?id=ED459671>
- Heesch, K. C., Masse, L. C., & Dunn, A. L. (2006). Using Rasch modeling to re-evaluate three scales related to physical activity: Enjoyment, perceived benefits and perceived barriers. *Health Education Research, 21*, 58–72. <http://dxdoi.org/10.1093/her/cyl054>
- İlhan, M., & Güler, N. (2016, September). *A Study with the Rasch model regarding the reverse directional item and number of response categories in Likert-type scales*. Paper presented at the 5th Congress of Measurement and Evaluation in Education and Psychology, Antalya, Turkey. Retrieved from http://epod2016.akdeniz.edu.tr/_dinamik/333/50.pdf
- İşeri, K., & Ünal E. (2012). Analyzing the Turkish teacher candidates' writing anxiety situations in terms of several variables. *Mersin University Journal of the Faculty of Education, 8*(2), 67–76.

- Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38(12), 1212–1218. <http://dxdoi.org/10.1111/j.1365-2929.2004.02012.x>
- Karakaya, İ., & Ülper, H. (2011). Developing a writing anxiety scale and examining writing anxiety based on various variables. *Educational Sciences: Theory & Practice*, 11, 691–707.
- Karatay, H. (2011). Süreç temelli yazma modelleri. Planlı yazma ve değerlendirme [Basic writing process models. Planned writing and evaluating]. In M. Özbay (Ed.), *Yazma eğitimi* [The education of writing] (pp. 21–43). Ankara, Turkey: Pegem Academy Publishing.
- Kuşdemir, Y., Şahin, D., & Bulut, P. (2016). Investigation of pre-service primary teachers' writing anxiety. *International Journal of Turkish Education Sciences*, 4(7), 100–109.
- Lee, M., Peterson, J. J., & Dixon, A. (2010). Rasch calibration of physical activity self-efficacy and social support scale for persons with intellectual disabilities. *Research in Developmental Disabilities*, 31(4), 903–913. <http://dxdoi.org/10.1016/j.ridd.2010.02.010>
- Linacre, J. M. (2012). *Many-facet Rasch measurement: Facets tutorial*. Retrieved from <http://www.winsteps.com/a/ftutorial2.pdf>
- Linacre, J. M. (2014). *A user's guide to FACETS Rasch-model computer programs*. Retrieved from <http://www.winsteps.com/a/facets-manual.pdf>
- Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment*, 1(1), 1–11. Retrieved from <http://files.eric.ed.gov/fulltext/ED506058.pdf>
- Penley, L. E., Alexander, E. R., Jernigan, I. E., & Henwood, C. I. (1991). Communication abilities of managers: The relationship to performance. *Journal of Management*, 17(1), 57–76. <http://dxdoi.org/10.1177/014920639101700105>
- Petzel, T. P., & Wenzel, M. U. (1993). *Development and initial evaluation of a measure of writing anxiety* (ERIC Document Reproduction Service No: ED 373072). Retrieved from <http://files.eric.ed.gov/fulltext/ED373072.pdf>
- Richmond, V. P., & Dickson-Markman, F. (1985). Validity of the writing apprehension test: Two studies. *Psychological Reports*, 56(1), 255–259. <http://dxdoi.org/10.2466/pr0.1985.56.1.255>
- Ricketts, S. N., Engelhard, G., & Chang, M. L. (2015). Development and validation of a scale to measure academic resilience in mathematics. *European Journal of Psychological Assessment*. <http://dxdoi.org/10.1027/1015-5759/a000274>
- Riffe, D., & Stacks, D. W. (1992). Student characteristics and writing apprehension. *Journalism & Mass Communication Educator*, 47(2), 39–49. <http://dxdoi.org/10.1177/107769589204700206>
- Shaver, J. P. (1990). Reliability and validity of measures of attitudes toward writing and toward writing with the computer. *Written Communication*, 7(3), 375–392. <http://dxdoi.org/10.1177/0741088390007003004>
- Singh, T. K. R., & Rajalingam, S. K. (2012). The relationship of writing apprehension level and self-efficacy beliefs on writing proficiency level among pre-university students. *English Language Teaching*, 5(7), 42–52. <http://dxdoi.org/10.5539/elt.v5n7p42>
- Smith, M. W. (1984). *Reducing writing apprehension*. Urbana, IL: National Council of Teacher of English. Retrieved from <http://files.eric.ed.gov/fulltext/ED243112.pdf>

- Stacks, D. W., Boozer, R. W., & Lally, T. D. P. (1983, March). *Syntactic language correlates of written communication apprehension*. Paper presented at the Annual Meeting of the Southwest Division of the American Business Communication Association (ERIC Document Reproduction Service No. ED226374). Retrived from <http://files.eric.ed.gov/fulltext/ED226374.pdf>
- Sudweeks, R. R., Reeve, S., & Bradshaw, W. S. (2005). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing*, 9(3), 239–261. <http://dxdoi.org/10.1016/j.asw.2004.11.001>
- Taşdelen Teker, G., Güler, N., & Kaya Uyanık, G. (2015). Comparing the effectiveness of SPSS and EduG using different designs for generalizability theory. *Educational Sciences: Theory & Practice*, 15, 635–645. <http://dxdoi.org/10.12738/estp.2015.3.2278>
- Temizkan, M. (2014). *Yaratıcı yazma süreci (Hikâye yazma)* [The creative writing process (Writing stories)]. Ankara, Turkey: Pegem Academy Publishing.
- Tennant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value in Health*, 7(1), 22–26. <http://dxdoi.org/10.1111/j.1524-4733.2004.7s106.x>
- Tezbaşaran, A. (1997). *Likert tipi ölçek geliştirme kılavuzu* [A guide to developing Likert-type scales]. Ankara, Turkey: Turkish Psychological Association.
- Tighe, M. A. (1987, March). *Reducing writing apprehension in English classes*. Paper presented at the Annual Meeting of the National Council of Teachers of English Spring Conference, Louisville, KY. Retrieved from <http://files.eric.ed.gov/fulltext/ED281196.pdf>
- Tiryaki, E. N. (2012). Determining writing anxieties of university students' from different variables. *Journal of Language and Literature Education*, 1(1), 14–21.
- Tok, M., & Potur, Ö. (2015). Trends of the academic studies in writing education area (2010–2014 years). *Journal of Mother Tongue Education*, 3(4), 1–25. Retrieved from <http://www.anadillegitimi.com/download/article-file/14934>
- Ungan, S. (2007). Devepment and importance of writing skills. *Journal of Erciyes University Social Sciences Institute*, 23(2), 461–462.
- Walker, E. R., Engelhard, G., & Thompson, N. J. (2012). Using Rasch measurement theory to assess three depression scales among adults with epilepsy. *Seizure*, 21(6), 437–443. <http://dxdoi.org/10.1016/j.seizure.2012.04.009>
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Yaman, H. (2010). Writing anxiety of Turkish students: Scale development and the working procedures in terms of various variables. *International Online Journal of Educational Sciences*, 2(1), 267–289.
- Yıldız, M., & Ceyhan, S. (2016). The investigation of 4th grade primary school students' reading and writing anxieties in terms of various variables. *Turkish Studies*, 11(2), 1301–1316. <http://dxdoi.org/10.7827/TurkishStudies.9370>
- You, H. S. (2016). Rasch validation of a measure of reform-oriented science teaching practices. *Journal of Science Teacher Education*, 27(4), 373–392. <http://dxdoi.org/10.1007/s10972-016-9466-3>

- Zanon, C., Hutz, C. S., Yoo, H. H., & Hambleton, R. K. (2016). An application of item response theory psychological test development. *Psicologia: Reflexao Critica*, 29(18), 1–10. <http://dxdoi.org/10.1186/s41155-016-0040-x>
- Zorbaz, K. Z. (2010). *İlköğretim okulu öğrencilerinin yazma kaygı ve tutukluğunun yazılı anlatım becerileriyle ilişkisi* [The relationship between middle school student's writing apprehension and blocking with their written expression skills] (Doctoral dissertation, Gazi University, Ankara, Turkey). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Zorbaz, K. Z. (2011). Writing apprehension and measurement of it. *e-Journal of New World Sciences Academy*, 6(3), 2271–2280.