International Journal of Assessment Tools in Education
2025, Vol. 12, No. 1, 93–112
https://doi.org/10.21449/ijate.1545220
journal homepage: https://dergipark.org.tr/en/pub/ijate **Research Article**

Turkish adaptation of the digital literacy scale: A rasch analysis

Hongwei Yang[1*], Müslim Alanoğlu[2], Songül Karabatak[2], Kelly D. Bradley[3]

[1]University of West Florida, United States
[2]Fırat University, Faculty of Education, Elazığ, Türkiye
[3]University of Kentucky, United States

ARTICLE HISTORY

Received: Sep. 07, 2024

Accepted: Nov. 12, 2024

Keywords:
Differential item functioning,
Digital literacy scale,
Scale validation,
Rasch analysis,
University students.

Abstract: The study took a Rasch measurement theory approach to validating the 10-item Digital Literacy Scale (DLS) using the unidimensional rating scale model (RSM). To that end, the study used the data from a sample of online Turkish university students. The study began the Rasch analysis with all 10 items in the scale and, to improve in the local independence assumption, identified and eliminated two items which did not adequately fit the RSM. Under the eight-item DLS, the assumptions of undimensionality and local independence were both satisfied and the fit of all individual items to the RSM was adequate. Next, the psychometric properties of the eight-item DLS were examined including rating scale effectiveness, relative endorsability of the items, differential item functioning (DIF) by each of three demographic variables: (a) gender, (b) connection device, and (c) grade level. Through the analysis, evidence of reliability and validity was identified which generally supports the use of the DLS instrument among the population of online Turkish university students from which the sample was obtained. The study also identified items which demonstrated either misfit to the model or DIF by the demographic variables, and recommends they be further reviewed and revised for future use.

1. INTRODUCTION

The term of digital literacy (DL) was first introduced and made known by Gilster (1997). This landmark book defined DL as the ability to comprehend and utilize information in multiple formats from various sources when the information is presented using computers. This definition, although it first emerged almost three decades ago, may still have relevance today cause it does not present any listing of specific digital skills or technologies which have evolved rapidly over the years. Instead, it approaches DL from a general and broad perspective to allow the interpretation and operationalization of the DL concept to easily develop as necessary (Ala-Mutka, 2011). In the research literature, digital literacy has had different definitions which could have substantial similarities and overlapping, could be based on different theoretical frameworks, and, with the emergence of new digital technologies, new tools, and new literacies, could evolve over

*CONTACT: Hongwei YANG pyang@uwf.edu University of West Florida, United States

e-ISSN: 2148-7456

time (Amin *et al*., 2022; Gillen & Barton, 2010; Olur & Ocak, 2021; Reddy *et al*., 2023; UNESCO, 2018).

1.1. A Multi-Literacy, Multi-Perspective Approach to Digital Literacy

Ng (2012a, 2012b) defined digital literacy as referring to the multitude of literacies related to the use of digital technologies which include both software and hardware employed by individuals for educational, social, and/or entertainment purposes both in schools and at home. Among such hardware and software are desktops, laptops, handheld devices (e.g., tablets), game consoles, smartphones, commercial and open-source programs, etc. Under this framework, digital literacy consists of cognitive, technical, and socio-emotional learning perspectives/dimensions overlapping between and among themselves, and involves the acquisition of skills under each of the three perspectives/dimensions in order to effectively engage with online/offline digital technologies.

In the education setting, academic digital literacy may be perceived as the ability and awareness to take advantage of digital technology as a learning tool and complete academic tasks in the right way, when also encompassing the cognitive, technical, and socio-emotional perspectives of the literacy (Anwar *et al*., 2023; Hwang *et al*., 2023). Academic digital literacy has an important role to play because it is viewed as the backbone of educational pedagogy (Anwar *et al*., 2023). Graduates with digital literacy competencies are likely to have substantially better job prospects because, with a vast majority of the jobs requiring digital literacy (Anthonysamy *et al*., 2020; Perera *et al*., 2016; Setiyowati & Razak, 2020), such competencies could well increase their productivity in the digital era.

1.2. Research Related to Digital Literacy

Over the years, particularly during the COVID-19 pandemic when there was a substantially increased exposure and use of digital platforms and technologies of all kinds in all walks of life, numerous studies have been conducted on a global scale which address digital literacy in a variety of contexts: (a) assessing the awareness and competencies of DL of individuals (e.g., students, teachers, etc.), (b) investigating how DL is related to other measures of interest (e.g., self-efficacy, self-esteem, professional competence, individuals' demographic variables, teachers' readiness to implement digital technologies), (c) examining the effectiveness of DL programs, (d) narrowing the DL skills gap, etc. (Aydınlar *et al*., 2024; Ceylan *et al*., 2023; Erol & Aydin, 2021; Garzon & Garzon, 2023; Liza & Andriyanti, 2020; Reddy *et al*., 2021; Reddy *et al*., 2023). A detailed review of DL-related studies is beyond the scope of the study. Readers are referred to related systematic reviews such as Nguyen and Habók (2024), Gutiérrez-Ángel *et al*. (2022), Wu *et al*. (2022), and Pangrazio *et al*. (2020).

Among the DL-related research are studies which address the development/adaptation and validation of the scales/assessment tools/instruments measuring digital literacy for various stakeholders, cultural contexts, etc. These studies use statistical and psychometric means to validate a multitude of instruments measuring DL.

Ng (2012a) presented one of the first instruments measuring digital literacy which is known as Digital Literacy Scale (DLS) in the literature (There are other DL instruments bearing the same name (e.g., that developed in Chandra *et al*. (2024)), but they are not discussed here in this study). Based on her DL framework, Ng developed this 10-item instrument and used it and several other instruments to investigate the learning of unfamiliar educational technologies among a group of Australian undergraduate students enrolled in an introduction course on eLearning. In her study, the DLS asked the students to evaluate their level of digital literacy using a 10-point Likert scale. Even though her study hardly investigated the psychometric properties of the DLS, many follow-up studies conducted psychometric validation of the instrument under various contexts (Anwar *et al*., 2023;

Hamutoğlu *et al*., 2017; Ustundag *et al*., 2017), or applied the DLS to content area research (Aydınlar *et al*., 2024; Durak & Seferoğlu, 2020; Erol & Aydin, 2021; Garzon & Garzon, 2023; Noorrizki *et al*., 2022; Tor *et al*., 2022).

1.3. Existing Validation Research of the DLS

The literature has witnessed multiple validation studies of the DLS instrument. A review of these studies is provided here in chronological order as the context justifying this new research.

Hamutoğlu *et al*. (2017) adapted a 17-item version of the DLS instrument into the Turkish context. Notably, Ng (2012a) presented DL as consisting of three dimensions / perspectives (i.e., cognitive, technical, and social-emotional) which were covered by 10 items in three subscales. Hamutoğlu *et al*. (2017) included those 10 items in their version of the DLS instrument and additionally treated the seven items measuring attitudes towards information and communications technology as the fourth dimension. This practice was not consistent with Ng (2012a) and several other studies, like Anwar *et al*. (2023), which were all conducted under a three-dimension structure for DL measured by 10 items. Based on their 17-item DLS, Hamutoğlu *et al*. (2017) first conducted an expoloratory factor analysis (EFA) using a sample of 185 students and next a confirmatory factor analysis (CFA) using a sample of 210 students. At the end of the analyses, they presented both validity (e.g., language validity) and reliability (e.g., Cronbach's α, test-retest reliability) evidence for the 17-item scale.

Ustundag *et al*. (2017) translated the 10-item version of the DLS instrument by Ng (2012a) into the Turkish context and administered the adapted instrument to a group of pre-service teachers studying science. Unlike Ng (2012a) who hardly investigated the psychometric properties of the original instrument, Ustundag *et al*. (2017) validated the adapted instrument using common statistical methods for scale validation. Among the analyses they conducted was an EFA which established that the DLS in Turkish was unidimensional and had relatively high internal consistency reliability.

Finally, Anwar *et al*. (2023) based their study on the digital literacy definition and the three- dimension DL model from Ng (2012a, 2012b). They adapted the 10-item DLS into the Indonesian context for university students to measure their academic digital literacy. In the validation of the adapted instrument, they primarily took the CFA approach using the data collected from a sample of 364 Indonesian students. Their final model included a second-order CFA model measuring academic digital literacy predicting the three dimensions of DL outlined by Ng (2012a, 2012b). Besides, they also reported several reliability statistics including Cronbach's α, composite reliability, and average variance extracted. Given the findings, they recommended the use of the adapted instrument among the Indonesian university students.

1.4. Research Gaps in the Existing DLS Instrument Validation Studies

Despite the multitude of existing DLS validation studies, in general, the validation of a scale should be a continuous process (Gocen & Sen, 2021; Nunnally, 1978). This process could require multiple validation iterations to continuously identify more evidence of an instrument's reliability and validity, and could also entail a broader variety of samples to further refine and validate the instrument under more research contexts. On the other hand, there is also room for improvement in the existing validation studies which warrants more research.

First, the existing studies primarily counted on the traditional EFA/CFA for continuous data without (any mention of) taking into consideration the typically ordinal, rating scale structure of the DLS item data. Even though treating ordinal data as continuous has been a long term debate (Frampton & Shepherd, 2011), the literature of multiple fields of studies (e.g., healthcare, nursing, etc.) has nevertheless indicated doing so could well run the risk of erroneous results and mis-inference (Adroher *et al*., 2018; Cape *et al*., 2010; Da Dalt *et al*., 2013, 2015; Hamilton & Chesworth, 2013; Miot, 2020).

Second, no existing validation studies have investigated whether the DLS instrument functioned equivalently across subgroups which may be of research interest (e.g., subgroups by participant demographic characteristics). Therefore, their findings did not address whether the DLS items were unfair to, for example, a particular gender subgroup. For instance, Erol and Aydin (2021) and Tor *et al*. (2022) each compared different gender (female vs. male) subgroups regarding the research participants' level of digital literacy measured by the DLS instrument.

Unfortunately, both studies did so without having first examined whether the DLS items were biased by gender, which left open the question on whether the statistically significant differences from the independent samples t tests they conducted regarding the measure of DL were artifacts of the characteristics of the biased items, if any, or due to variations of participants' digital literacy at the scale and the subscale levels. Besides gender, the literature has indicated that digital literacy could be impacted by multiple demographic factors which include, but are not limited to, age, education, family income, use of smartphones and the Internet, years of service in the profession, daily Internet usage time, technology usage level, social media usage in distance education (Erol & Aydin, 2021; Noorrizki *et al*., 2022; Tor *et al*., 2022; Urbancikova *et al*., 2017). In order to examine the difference in the DLS scores, if any, across the subgroups specified by a demographic variable, the DLS items should be first verified to function the same way across these subgroups. This topic has not been investigated in the existing studies validating the DLS instrument.

1.5. Rasch Analysis as an Instrument Validation Tool

Rasch Measurement Theory (RMT) is a latent modeling framework which is based on modern test theory. In Rasch analysis, the raw, ordinal data (e.g., responses to Likert type items like DLS items) of the instrument are transformed to interval/continuous measures of participant ability and item difficulty on a logit scale along which a side-by-side comparison of participants and items is made (Andrich & Marais, 2019; Bond & Fox, 2015). Many, but not all, Rasch models assume that Rasch measurement involves a single, underlying construct (i.e., assumption of unidimensionality) either increasing or decreasing monotonically along the interval logit scale. Under the RMT, to make valid comparisons across different subgroups regarding a latent construct (e.g., digital literacy), the items should function the same way across different subgroups of participant demographic characteristics (e.g., gender) (Hagquist *et al*., 2009; Hagquist, 2019). Otherwise, comparisons of scores across the subgroup participant characteristics (e.g., female vs. male) may be invalid. Such a violation of the requirement of invariance across subgroups is known as differential item functioning (DIF; Hagquist, 2019). In summary, RMT methods are designed to properly handle the ordinal categorical data. They can complement the traditional methods in psychometrics (e.g., proportion of correct responses as a measure of item difficulty) to provide additional evidence of reliability and validity of an instrument. Over the years, they have been widely used in studies (e.g., those validating a scale) in education including those of online education (e.g., Ningsih *et al*. (2021)), artificial intelligence in education (e.g., Capinding (2024)), among others.

1.6. Research Questions

Rasch analysis provides a detailed analysis of many aspects of an instrument when also being able to address the research gaps (e.g., taking into consideration the ordinal, rating scale structure of the DLS data, investigating item DIF, etc.) outlined above.

However, an extensive literature review indicates that there have not been any studies reporting the psychometric properties of the DLS instrument by means of RMT.

Given the discussions above, the study proposed three research questions (RQs) regarding the DLS instrument:

1. RQ1: Does the DLS instrument measure a unidimensional construct of digital literacy?
2. RQ2: What are the psychometric properties of the DLS instrument, after properly taking into account the rating scale structure of the DLS response data?
3. RQ3: Do the DLS items function equivalently across the subgroups specified by participants' demographic measures?

1.7. Organization of Research

The study is organized as follows. The study begins with an introduction of the research context, which is followed by a review of the existing DLS scale validation research and gaps in such research. Rasch analysis is introduced as a psychometric method addressing the gaps. Next come the research questions with regard to the DLS instrument which were formulated based on the literature review, outlined research gaps, and introduction of Rasch analysis. The study proceeds to a methodology section which examines the psychometric properties of the DLS under Rasch analysis. In the end, the study discusses the findings, implications, and limitations and future research before providing the final conclusions.

2. METHOD

2.1. DLS Instrument and Demographic Measures

This study used the 10-item (Table 1) version of the DLS instrument by Ng (2012a). Each DLS item is measured on a five-point Likert scale, ranging from 1 = *Strongly Disagree* to 5 = *Strongly Agree*. Note that, although Ng (2012a) developed the DLS on a 10-point Likert scale, many follow-up (scale validation or content area) studies (e.g., Ustundag *et al*. (2017), Garzon and Garzon (2023), among others) used a five-point Likert scale, instead, and this study followed the same practice. Finally, because all 10 DLS items are positively worded, a higher score on an individual DLS item, a subscale, and the scale as a whole corresponds to a higher level of digital literacy.

Table 1. *DLS items*.

Items	Item statements
DLS01	I know how to solve my own technical problems.
DLS02	I can learn new technologies easily.
DLS03	I keep up with important new technologies.
DLS04	I know about a lot of different technologies.
DLS05	I have the technical skills I need to use ICT[a] for learning and to create artifacts (e.g., presentations, digital stories, wikis, blogs) that demonstrate my understanding of what I have learned.
DLS06	I have good ICT[a] skills.
DLS07	I am confident with my search and evaluation skills in regards to obtaining information from the Web.
DLS08	I am familiar with issues related to web-based activities e.g., cyber safety, search issues, plagiarism.
DLS09	ICT enables me to collaborate better with my peers on project work and other learning activities.
DLS10	I frequently obtain help with my university work from my friends over the Internet e.g. through Skype, Facebook, Blogs.

Note. The sample size is consistently $n = 404$ across all 10 DLS items.
[a]ICT = Information and Communication Technology.

Regarding the demographic items, there were three dichotomously-coded ones: (a) gender, (b) connection device, and (c) grade level. Gender consists of the two categories of females and males, connection device the two categories of computers (desktop and laptop) and handheld devices (smart phone and tablet), and grade level the two categories of lower (first- and second- years) and higher (third- and fourth-years) grades of undergraduate students.

2.2. Participants and Data Collection

After securing the required approval from the research ethics committee of the research site of a Turkish university, the study proceeded to obtain a convenience sample. The data were collected in the university as part of a larger cross-sectional study among its undergraduate students of education taking online courses. After properly preparing the collected data, the final sample size of each item was consistently n = 404.

2.3. Rasch Analysis

The data were first summarized using descriptive statistics which were based on several breakdowns of the participants' demographic characteristics. Next, a Rasch analysis of the data was conducted using the Rasch Rating Scale Model (RSM) in Winsteps 5.6.4.0 (Linacre, 2023). An RSM is a type of Rasch model for polytomous data usually produced from a Likert scale.

The model requires every item should have the same number of response categories (e.g., the DLS instrument where all items have five response options). Besides, to each item, the model applies the same number of response thresholds, with which to progress from one response option to the next (e.g., from *Agree* to *Strongly Agree*); across all items, the relative distance between each pair of threholds remains the same, although each item is still allowed to have its own level of difficulty.

The RSM-based Rasch analysis began with all 10 items in the model and assessed the statistical assumptions (i.e., assumptions of unidimensionality and local independence) underlying the RSM and the fit of the data to the model. In the case of a problem (e.g., assumption violation, inadequate fit of the item data to the model, etc.), appropriate measures were taken to address it. After the assumptions were fully satisfied and the fit of the item data to the model was improved to an acceptable level, the Rasch analysis of the instrument was advanced to produce more evidence of reliability and validity.

3. RESULTS

As was shown in Table 1, the dataset contained 404 participants providing complete responses to all 10 DLS items. Therefore, the dataset led to a high participant-item ratio of about 40:1, satisfying the criterion that the sample size should be at least six times the number of items for stable results in factor analysis of which Rasch analysis is a special type for categorical data (Bartholomew *et al.*, 2008; Mundfrom *et al.*, 2005; Skrondal & Rabe-Hesketh, 2004).

3.1. Descriptive Statistics

Regarding the participant demographics, the sample of 404 participants ranged from 18 to 46 years old in age (M = 24.03, SD = 4.39) and consisted of 308 females and 96 males. They used different devices to connect to the Internet: (a) n = 21 using a desktop, (b) n = 156 using a laptop, (c) n = 216 using a smart phone, and (d) n = 8 using a tablet. Finally, they came from four different grades: (a) n = 31 from first-year, (b) n = 53 from second-year, (c) n = 40 from third-year, and (d) n = 280 from fourth-year.

Further, the mean response scores for individual items (computed by averaging all responses to each item across all participants who responded to the item) fall between *Agree* (= 4) and *Neither Agree nor Disagree* (= 3), ranging from 3.11 for DLS06 and to 3.89 for

DLS07. All items put together, the most frequently selected category is *Agree* (32.4%), which is immediately followed by *Neither Agree nor Disagree* (28.2%).

Finally, Table 2 documents the response frequencies of the categories of individual DLS items. According to the table, *Agree* is the most frequently selected category on five items (ranging from 27.7% for DLS10 to 40.6% for DLS07), and *Neither Agree nor Disagree* is most frequently selected on the other five items (ranging from 31.2% for DLS05 to 35.1% for DLS01). As a summary, the observations from descriptive statistics suggest the student participants mostly perceived neutrally to favorably of how well the items described their levels of digital literacy.

Table 2. *Summary of responses to all 10 DLS items*.

Items	Strongly Disagree (%)	Disagree (%)	Neither Agree nor Disagree (%)	Agree (%)	Strongly Agree (%)
DLS01	4.7	10.9	35.1	32.4	16.8
DLS02	3.2	6.9	19.8	39.9	30.2
DLS03	5.7	10.1	24.3	35.9	24.0
DLS04	5.0	16.8	31.9	28.5	17.8
DLS05	5.9	16.1	31.2	30.0	16.8
DLS06	10.4	18.8	33.7	23.5	13.6
DLS07	1.7	6.9	21.5	40.6	29.2
DLS08	5.2	12.4	31.4	29.0	22.0
DLS09	3.0	8.2	25.5	36.1	27.2
DLS10	8.4	14.6	27.5	27.7	21.8

3.2. Rasch Analysis

The study began with all 10 DLS items analyzed under the RSM and assessed whether the two statistical assumptions of the RSM were satisfied: unidimensionality and local independence (Bond & Fox, 2015).

3.2.1. *Analyzing 10-item DLS*

3.2.1.1. Assumption of Unidimensionality. This assessment of the unidimensionality assumption served to see if the DLS instrument, as a whole, measures a single underlying construct of digital literacy that the instrument was designed to measure. To that end, a principal component analysis (PCA) was used of the correlation matrix of standardized Rasch residuals (Bond & Fox, 2015; Linacre, 2023).

According to the Winsteps PCA output, the statistics of explained raw score variance in the observations/observed data by measures (i.e., items and persons) in the *Observed* column and those in the *Expected* column were about the same size (for persons, 46.6% under *Observed* vs. 46.7% under *Expected*; for items, 9.1% under *Observed* vs. 9.1% under *Expected*), indicating there was no problem in the model estimation and that the data provided an adequate fit to the Rasch model assuming unidimensionality (Linacre, n.d.; Linacre, 2018, September 2). Second, the contrasts were examined which were computed after the Rasch dimension was extracted from the data. Specifically, the first contrast (i.e., the first dimension beyond the Rasch dimension) had an eigenvalue of 1.9717, which was lower than 2, the size of an eigenvalue expected by chance. This evidence did not support the existence of a secondary dimension in the data (Linacre, 2023). Based on the multiple pieces of evidence from both the statistical analyses and the literature, the study concludes with the unidimensionality (i.e., Rasch dimension) of the 10-item DLS.

3.2.1.2. Assumption of Local Independence. Also assessed here was the local independence assumption which states that, after controlling for the underlying latent trait of digital literacy, the responses to one survey item do not covary with the responses to other items (Aryadoust *et al*., 2021; Borsboom, 2005). That is, in Rasch measurement, since DLS items are regressed on the latent variable of digital literacy, the local independence assumption requires that the unexplained variances in the DLS items should not correlate with each other. For the 10-item DLS, the local independence assumption was assessed using the correlations between the residuals of the DLS items (i.e., Q3 coefficients). (Fan & Bond, 2019; Lee, 2004; Wright, 1996; Yen, 1984). A Q3 coefficient larger than .30 in absolute value indicates a respectable degree of local dependence. Examining the Winsteps output of the largest standardized residual correlations of DLS items showed that the correlations in absolute value between the residuals of three pairs of items were higher than .30: (a) (-.32) between DLS03 and DLS10, (b) (-.31) between DLS02 and DLS06, and (c) (-.30) between DLS04 and DLS10. Therefore, there was a violation of the assumption of local independence among the three pairs of items. To find more evidence for addressing this assumption violation, individual item fit was next examined.

3.2.1.3. Individual Item Fit. Examining the item fit output containing the mean-square (MNSQ) infit and outfit statistics, one and only one item, DLS10, had an unusually large infit MNSQ (1.84) and outfit MNSQ (1.95) at the same time. Because these statistics were greater than 1.50, it indicates that, with this item, off-variable noise was markedly greater than useful information. As a result, even though these diagnostic statistics were (close to but) still not higher than the reshold of 2.00 indicative of degradation of measurement, the item may nonetheless need to be further scrutinized and revised to remedy its misfit to the model. Other than DLS10, the other nine items were all productive of measurement. None of them exhibited any substantial misfit to the Rasch model because their infit and outfit statistics were at most 1.28 (infit MNSQ) and 1.29 (outfit MNSQ) for DLS06 and at least 0.67 (infit MNSQ) and 0.73 (outfit MNSQ) for DLS04, which all fell into the range of 0.50 – 1.50 indicating productive of measurement. Finally, the point-polyserial correlations for all 10 items were high and positive where the lowest correlation was that for DLS10 at .60 and all other correlations were at least .71 (DLS06 and DLS07), indicating the orientation of the scoring on each DLS item was well aligned with the orientation of the latent variable measured by this instrument and that the items had adequate discriminatory power. The point-polyserial correlation for DLS10 was positive but was also markedly lower than the other nine correlations. Therefore, DLS10 probably did not have as much discriminatory power as any of the other nine items (Bond & Fox, 2015; Linacre, 2023).

3.2.2. *Analyzing eight-item DLS*

Based on the analyses above, DLS10 was identified as not having an adequate fit to the model and was among the items which led to a violation of the local independence assumption. Therefore, DLS10 was removed and the above analyses were repeated with the remaining nine items. This time, the unidimensionality assumption continued to be satisfied. But, there was still one pair of items, DLS02 and DLS06, with the Q3 coefficient being (-.34) whose absolute value was higher than .30. Therefore, with the remaining nine items, the local independence assumption was violated again. Next, the fit of individual items was examined. Among the remaining nine items, DLS06 had the highest outfit MNSQ (1.53) which did not fall into the range of 0.50-1.50, and its infit MNSQ (1.48), although also the highest, fell into the range. All other items had both infit and outfit MNSQ statistics in the range of 0.50-1.50. Given the information above, out of the only pair of items (DLS02 and DLS06) whose Q3 coefficient indicated a violation of the local independence assumption, DLS06 was removed from further consideration. There was a total of eight items left in the scale.

3.2.2.1. Overview of the Eight-Item Scale. Next, the eight-item scale was examined under the Rasch rating scale model. First, the assumption of unidimensionality was satisfied. The statistics of explained raw score variance in the *Observed* column and those in the *Expected* column were virtually identical (for persons, 53.7% under *Observed* vs. 53.6% under *Expected*; for items, 8.1% under *Observed* vs. 8.1% under *Expected*), indicating there was no problem in model estimation and the data provided an adequate fit to the unidimensional Rasch model. The first contrast beyond the Rasch dimension had an eigenvalue of 1.7880, which was lower than 2 and thus did not support the existence of a secondary dimension in the data. Second, the local independence assumption was satisfied under the eight-item scale because, among the correlations between the residuals of the eight DLS items, the highest was .29 in absolute value which was lower than .30. Third, regarding the fit of individual items to the model, all infit and outfit MNSQ statistics fell into the range of .50-1.50 (highest and lowest infit MNSQ statistics were, respectively, 1.20 for DLS08 and 0.78 for DLS04; highest and lowest outfit MNSQ statistics were, respectively, 1.20 for DLS08 and 0.76 for DLS02). Therefore, given the statistics above, the eight-item scale met the assumptions of undimensionality and local independence and provided an adequate fit at both the overall and individual item levels. It was therefore further examined and interpreted under the Rasch model.

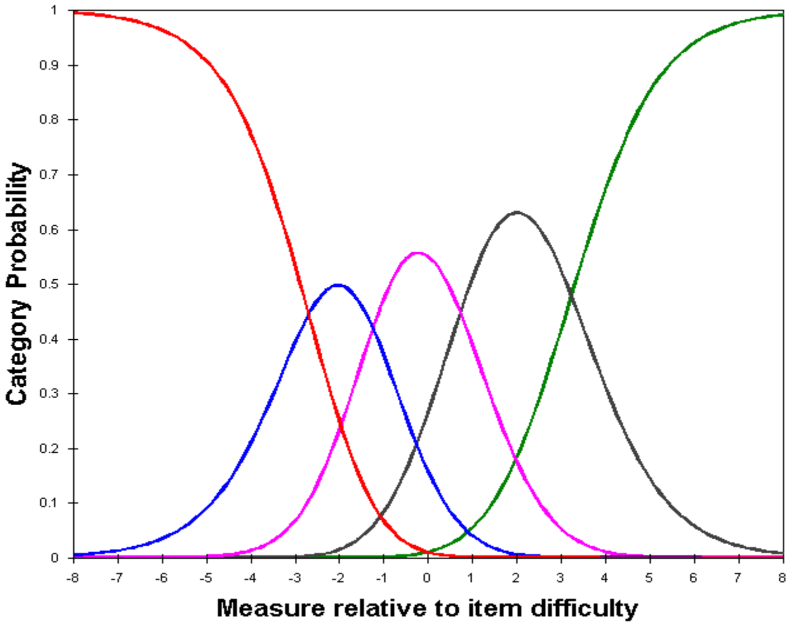
3.2.2.2. Separation and Reliability. In the eight-item DLS, person and item separation statistics were, respectively, as high as 2.71 and 6.16. The high person separation statistic indicated the DLS instrument was adequately sensitive to distinguish between individual participants with higher and lower levels of digital literacy, and the high item separation statistic indicated the sample was large enough to confirm item difficulty/endorsability/agreeability hierarchy. Regarding the reliability statistics, person reliability was .88 (i.e., the DLS instrument discriminated the participants into adequate levels of digital literacy), and item reliability was also very high at .97 (i.e., the sample was large enough to precisely locate the items on the underlying latent difficulty/endorsability/agreeability continuum) (Bond & Fox, 2015; Linacre, 2023).

3.2.2.3. Rating Scale Effectiveness for DLS. The study also examined the rating scale effectiveness of the eight items in DLS. First, according to the response category probability curves shared by all eight items in the scale (Figure 1), each category had a distinctive peak indicating it was a meaningful endorsement choice for the participants at a certain level of ability as measured in DLS. Stated differently, the Turkish student participants were capable of adequately separating one response category from another in the eight DLS items, which served as evidence of validity (Bond & Fox, 2015; Linacre, 2023).

Second, regarding the quality of the rating scale categories, none of the outfit MNSQ statistics on the five categories was greater than 2. The infit MNSQ statistics ranged from 0.86 for *Agree* (= 4) to 1.14 for *Strongly Agree* (= 5) and the outfit MNSQ statistics from .85 for *Agree* (= 4) to 1.14 for *Disagree* (= 2), indicating that none of the categories was introducing more noise than meaning into the measurement process and thus warranted further empirical investigation (e.g., considered as a candidate for collapsing with adjacent categories) (Bond & Fox, 2015). Third, the measure of Andrich threshold advanced in a stepwise manner (the four threshold statistics (-2.70 < -1.25 < 0.72 < 3.23) ascended monotonically in value up the rating scale) as anticipated, indicating that the lower threshold was always smaller than the higher threshold in an adjacent pair of categories. Stated differently, there was no disordering of thresholds (Bond & Fox, 2015; Linacre, 2023).

As a summary, the findings here support the rating scale structure of the DLS instrument functioned in the intended way, and that the response categories were correctly and consistently interpreted by the student participants as the sequence of most likely outcomes.

Figure 1. *Response category probability curves shared by all eight items in the DLS instrument.*



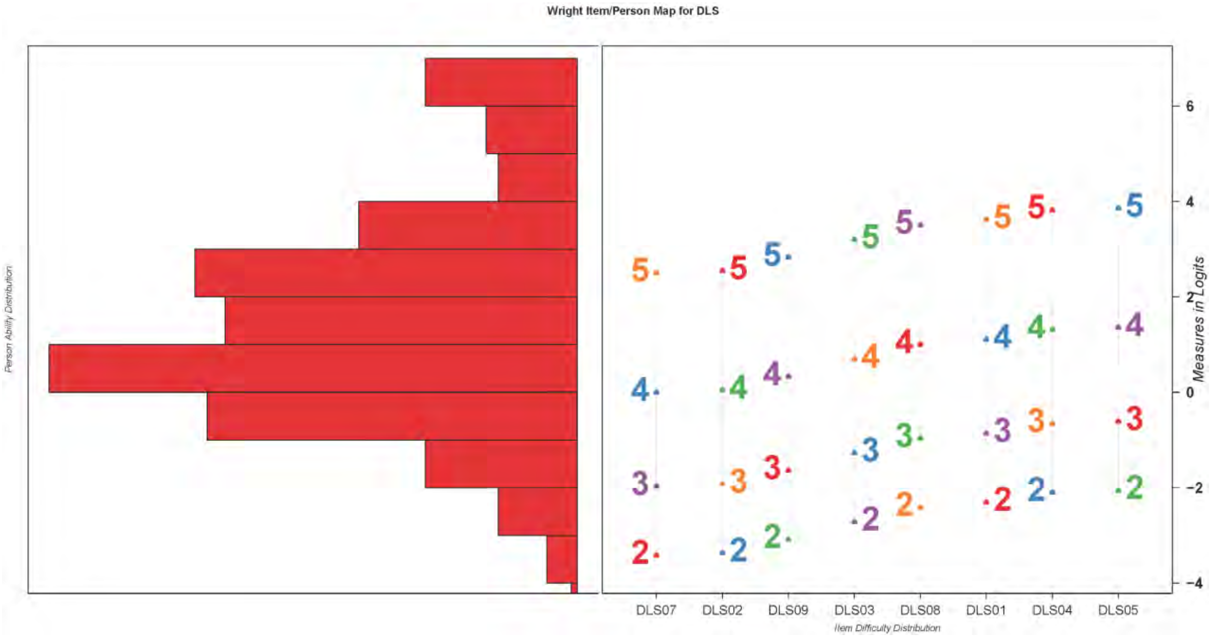
Note. Curve peaks for response categories (from left to right): 1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Neither Agree nor Disagree*, 4 = *Agree*, 5 = *Strongly Agree*.

3.2.2.4. Wright Item/Person Map for DLS. The Wright, item/person map in Figure 2 visually demonstrates and rank-orders the relative difficulty/endorsability/agreeability of the eight DLS items and students' level of digital literacy. In the right portion of the panel, from left to right, items are ranked from the most favorite item (i.e., easiest to endorse) to the least favorite item (i.e., hardest to endorse) and the four Andrich thresholds (i.e., step values) of the RSM for each individual DLS item with five response categories are indicated vertically and in ascending order by numeric values of 2, 3, 4, and 5 above that item; in the left portion of the panel, from bottom to top, students are ranked from those who had the lowest level of digital literacy to those who had the highest level of digital literacy (Linacre, 2023).

Based on Figure 2, the student participants most easily endorsed DLS07, "*I am confident with my search and evaluation skills in regards to obtaining information from the Web.*" and DLS02, "*I can learn new technologies easily.*". Next, in an ascending order of difficulty, the students almost equally easily endorsed DLS09, "*ICT enables me to collaborate better with my peers on project work and other learning activities.*". However, when it came to DLS03, "*I keep up with important new technologies.*", the item was more difficult to endorse by the students than the previous items. Next, at a higher level of difficulty was DLS08, "*I am familiar with issues related to web-based activities e.g., cyber safety, search issues, plagiarism.*". Even more difficult to endorse was DLS01, "*I know how to solve my own technical problems.*". Finally, the two most difficult-to-endorse items were DLS04, "*I know about a lot of different technologies.*", and, subsequently, DLS05, "*I have the technical skills I need to use ICT for learning and to create artifacts (e.g., presentations, digital stories, wikis, blogs) that demonstrate my understanding of what I have learned.*".

The results indicated that, overall, the student participants willingly demonstrated their confidence in the level of digital literacy. However, that confidence might not have easily translated into the participants' actual digital literacy. Therefore, it was not surprising to see that they were hesitant to acknowledge that they actually had the knowledge, technologies, or skills.

Figure 2. *Wright item/person map for validating DLS.*



Note. In the right portion of the panel, the four Andrich thresholds (i.e., step values) of the RSM for each individual DLS item with five response categories are indicated vertically and in ascending order by numeric values of 2, 3, 4, and 5 above that item.

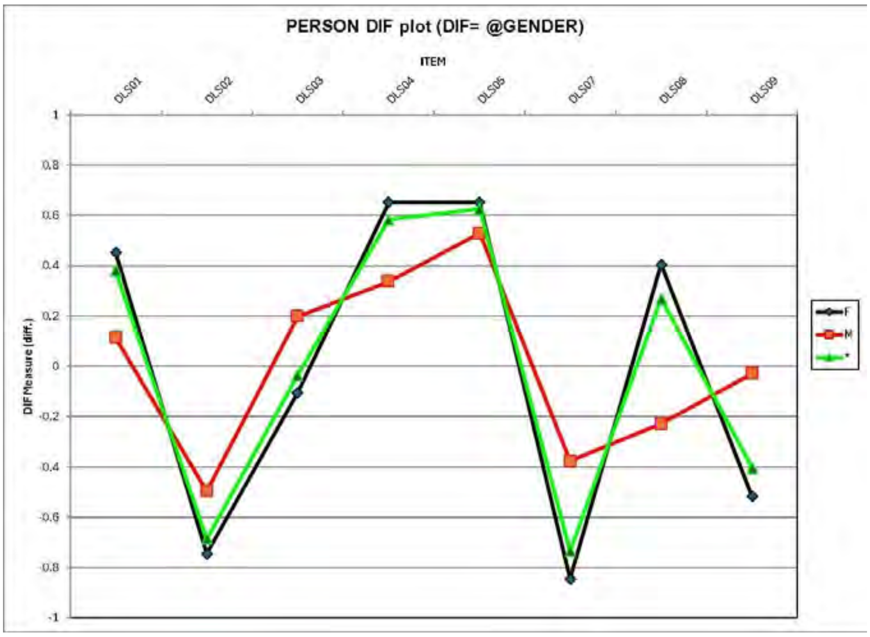
3.2.2.5. Differential Item Functioning Analysis of DLS. A pairwise differential item functioning analysis of the items in DLS by each of three dichotomously-coded demographic items (i.e., gender, connection device, and grade level) was conducted where the null hypothesis was set up that each DLS item had the same level of difficulty for the two subgroups specified by each demographic variable. Both statistical significance and substantive significance were assessed using, respectively, (a) the Rasch-Welch t and the Mantel χ^2 tests and (b) the cumulative log-odds ratio approximating the DIF size for polytomous data (Linacre, 2023). The results of the three DIF analyses are outlined in Table 3.

Table 3. *Results of three DIF analyses.*

Items	Female Minus male			Computer Minus handheld device			Higher grade Minus lower grade		
	Rasch-Welch *t* test	Mantel χ^2 test	DIF size	Rasch-Welch *t* test	Mantel χ^2 test	DIF size	Rasch-Welch *t* test	Mantel χ^2 test	DIF size
DLS01	.0769	.0898	.48	.5284	.7322	.08	.9100	.9278	-.02
DLS02	.2090	.0821	-.53	.4600	.1165	.41	.9007	.9874	.00
DLS03	.1120	.1179	-.45	.4778	.1850	.34	1.0000	.6525	.13
DLS04	.0965	.1176	.43	1.0000	.9408	-.02	.4327	.2969	-.30
DLS05	.5103	.4180	.21	1.0000	.8805	-.03	1.0000	.6130	-.13
DLS07	.0189	.0401	-.62	1.0000	.6507	.11	.9089	.9262	.03
DLS08	.0013	.0133	.75	.6728	.3403	-.23	.1289	.2401	.33
DLS09	.0123	.1333	-.41	.1582	.0528	-.45	.4965	.7500	-.09

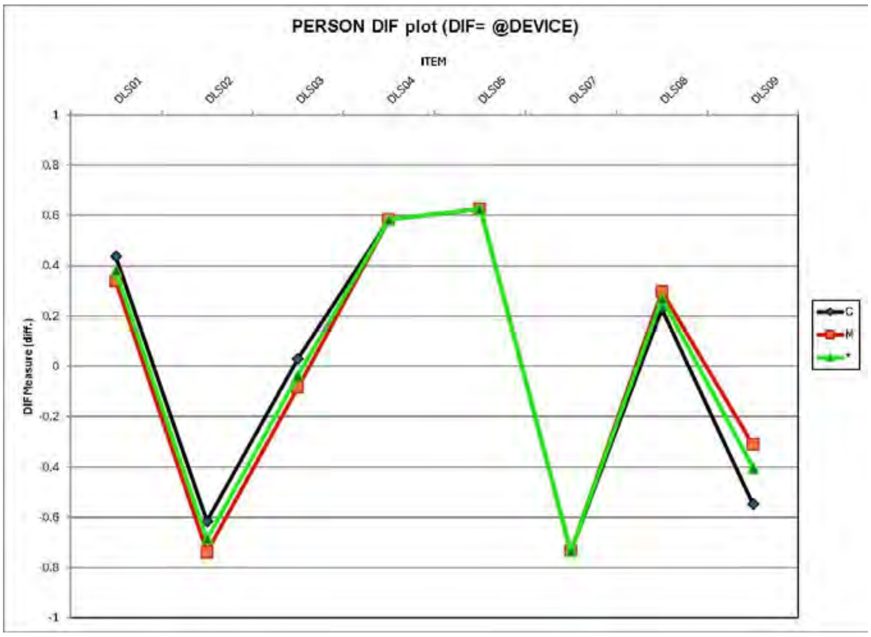
3.2.2.5.1. DIF analysis by Gender. Per the measure of DIF contrast for each item computed as the difficulty estimate of the item for females minus that for males, two items were statistically significant at the .05 level of significance on both the Rasch-Welch t and the Mantel χ^2 tests: DLS07 and DLS08. DLS07 had a negative DIF contrast and therefore was easier for the female subgroup than for the male subgroup. In comparison, since DLS08 demonstrated a positive DIF contrast, this item was the other way around (i.e., more difficult for the female subgroup than for the male subgroup). Next, both items demonstrated a moderate to large level of DIF (i.e., Level C DIF): (a) DLS07 with a DIF size of (-0.62) and (b) DLS08 with a DIF size of 0.75. Finally, DLS09 was significant on the Rasch-Welch t test, p = .0123, but not significant on the Mantel χ^2 test, p = .1333. It had a negligible DIF size of (-0.41) (i.e., Level A DIF) (Linacre, 2023; Zwick, 2012; Zwick *et al*., 1999). Finally, the DIF analysis by gender is presented graphically in Figure 3.

Figure 3. *DIF analysis by gender.*



3.2.2.5.2. DIF analysis by Connection Device. Per the measure of DIF contrast for each item computed as the difficulty estimate of the item for computers minus that for handheld devices, none of the eight items was statistically significant at the .05 level of significance on any of the Rasch-Welch t and the Mantel χ^2 tests. Next, all but one item demonstrated a negligible level of DIF (Level A DIF) and DLS09 demonstrated a slight to moderate level of DIF of (-.45) (Level B DIF) (Linacre, 2023; Zwick, 2012; Zwick *et al*., 1999). Finally, the DIF analysis by connection device is presented graphically in Figure 4.

Figure 4. *DIF analysis by connection device.*



3.2.2.5.3. DIF analysis by Grade Level. Per the measure of DIF contrast for each item computed as the difficulty estimate of the item for higher grade students minus that for lower grade students, none of the eight items was statistically significant at the .05 level of significance on any of the Rasch-Welch t and the Mantel χ^2 tests. Next, all items demonstrated a negligible level of DIF (Level A DIF) (Linacre, 2023; Zwick, 2012; Zwick *et al*., 1999). Finally, the DIF analysis by grade level is presented graphically in Figure 5.

Figure 5. *DIF analysis by grade level.*

4. DISCUSSION

In the digital era, digital literacy is constantly referred to and its importance is evidenced by the numerous efforts at different levels (e.g., regional, national, etc.) to develop and implement DL frameworks and strategic plans to support and improve their citizens' digital literacy (UNESCO, 2018). In an effort to contribute to the proper measurement of digital literacy, the study adapted to the Turkish context the Digital Literacy Scale through a Rasch measurement theory perspective. The study identified evidence of undimensionality of the eight-item DLS instrument, which is very close to the conclusion of Ustundag *et al*. (2017) stating that all ten DLS items constituted a unidimensional measure of digital literacy. By contrast, the study may differ from other DLS adaptation research, such as Anwar *et al*. (2023) and Hamutoğlu *et al*. (2017), in terms of the conclusion on scale dimensionality and an attempt is made at a later point in this study to address such a discrepancy. Besides, new evidence of reliability and validity was found in the study which provided more insights into the psychometric properties of the DLS items. Three research questions were proposed and addressed.

4.1. Addressing Research Questions

Regarding RQ1, the study found that, with all 10 items in the scale, the fundamental unidimensionality assumption of the RSM was satisfied. However, the 10-item scale led to a violation of the local independence assumption. After identifying and removing two items, DLS10 and DLS06, the unidimensionality and local independence assumptions were both satisfied under the eight-item DLS. Finally, because DLS10 and DLS06 exhibited a misfit to the model, DLS10 in particular, they merit further review and revision to prevent them from degrading the measurement of digital literacy.

Regarding RQ2, the study conducted a Rasch analysis under the eight-item DLS to investigate item/person separation and reliability, rating scale effectiveness, and relative endorsability of items. In the analysis, measures of item/person separation and reliability were all high. The high level of person separation indicated the DLS instrument was able to distinguish between participants with higher and lower levels of digital literacy, and the high level of item separation indicated the sample was adequately large to confirm item endorsability hierarchy. That the item and person reliability measures were high suggested the item difficulty and participant ability measures would be highly reproducible, should the same test be administered to the same group of student participants repeatedly. Next, a diagnostic analysis of the rating scale effectiveness in the eight-item scale indicated its response categories functioned as intended, and that the participants were able to adequately separate one response category from another and correctly and consistently interpret the response categories. Finally, in the Wright, item/person map, the item hierarchy measuring relative endorsability was demonstrated. Overall, the student participants easily agreed they were confident in their level of digital literacy, but that confidence did not easily translate into the actual digital literacy skills they would acknowledge they had.

Regarding RQ3, the study conducted a DIF analysis under the eight-item DLS to see if any items were endorsed to different extents by the two subgroups specified by each of the three demographic variables: (a) gender, (b) connection device, and (c) grade level. First, under gender, two items, DLS07 and DLS08, demonstrated statistical significance as measured by both the Rasch-Welch t and the Mantel χ^2 tests. DLS07 was easier for females to endorse than for males, whereas the DIF of DLS08 was in the opposite direction. Both DLS07 and DLS08 demonstrated a Level C DIF. Besides DLS07 and DLS08, DLS09 was significant on the Rasch-Welch t test only and demonstrated a Level A (i.e., negligible) DIF. Second, under connection device, none of the eight items was significant on any of the Rasch-Welch t and the Mantel χ^2 tests. DLS09 was the only item which demonstrated a slight to moderate level of (i.e., Level B) DIF. Third, under grade level, none of the eight items was significant on any of the Rasch-Welch t and the Mantel χ^2 tests, neither was there any item demonstrating a level of DIF beyond negligible.

Because several items were flagged as having gender-related DIF in this study, it is reasonable to be wondering if the gender-based comparisons presented in studies like Erol and Aydin (2021) would have led to different results. Therefore, such studies should probably have begun with an assessment of whether gender-related DIF existed on any items before comparing the two gender subgroups on digital literacy at the scale and subscale levels. This assessment is necessary because differences in DLS scores between the gender subgroups could reflect the characteristics of DLS items instead of variations in the participants' level of digital literacy that the study intended to assess. In the long run, it is important to be aware of any bias coming from item DIF, particularly if thresholds are to be applied to the DLS scores to inform decisions on diagnosis and subsequent interventions or treatments. When DIF exists, the associated bias could lead to under- or over-intervention or treatment for certain subgroups, depending on the direction of the bias. Accordingly, it is important for the DLS instrument to be assessed for DIF and the extent to which it exists should be taken into consideration when interpreting the DLS scores (Cameron *et al*., 2014).

4.2. Implications

The DLS instrument, together with its adaptation using the Rasch measurement theory in this study, has implications for assisting researchers, policymakers, instructional designers, and online instructors. This instrument is well-suited for gaining insights into the specific digital literacy requirements of Turkish university students as they engage with digital technologies. Furthermore, it may also serve as a catalyst for targeted interventions and programs aimed to improve the digital literacy skills of Turkish university students. By properly measuring the digital literacy of university students, this assessment tool likely has the potential to improve efficiency, effectiveness, and success in the adoption of ICT-based online education practices.

4.3. Limitations and Future Research

This study has its limitations which may serve as grounds of future research. First, this study is limited to examining the effects of three demographic variables as potential sources/covariates of DIF and the findings already cast doubts on the results from the existing literature (e.g., Erol and Aydin (2021)). Future research might investigate other possible sources/covariates (e.g., race/ethnicity) of DIF which might be of interest to content area researchers. Second, in this study, the DLS survey was not completed over time and no consideration was given to the ability of the DLS instrument to identify changes in digital literacy longitudinally. Future research might focus on the longitudinal measurement invariance aspect of the psychometric properties of the DLS instrument to see whether the DLS items assess the same digital literacy construct invariantly across time (Horn & McArdle, 1992; Liu *et al*., 2017; Meredith, 1993). For example, yearly, as in Lazonder *et al*. (2020).

Third/finally, the current study is limited in that it did not evaluate the bifactor model as an alternative structural representation (e.g., dimensionality) of the DLS instrument. Although this study presented evidence of undimensionality and this conclusion is largely consistent with that from certain previous research (e.g., Ustundag *et al*. (2017)), there are nonetheless other DLS validation studies (e.g., Ng (2012a) and Anwar *et al*. (2023)) which demonstrate the DLS instrument is multi-dimensional. A tentative explanation for this discrepancy might be that neither conclusion adequately explains the true dimensionality of the DLS instrument. Instead, a combination of the two solutions in the form of a bifactor model (Chen *et al*., 2012; Gignac & Kretzschmar, 2017; Reise, 2012; Reise *et al*., 2007; Rodriguez *et al*., 2016a, 2016b) might provide a fuller representation of the underlying structure of the DLS instrument. Under a bifactor model, previous research has indicated that an instrument consisting of multiple dimensions/subscales could be consistent with both a unidimensional and a multi-dimensional model but may be alternatively and likely better represented by the bifactor structure (Reise *et al*., 2007). For example, the bifactor structure might be able to more effectively handle the violation of the local independence assumption due to item clustering demonstrated earlier in the study. Besides, the DL framework proposed in Figure 1 from Ng (2012a) features three separate circles (e.g., representing the three dimensions of DL: cognitive, technical, and socio-emotional learning) overlapping in pairs and in an intersection of all three circles. The bifactor model can not only include the overlap of each pair of DL dimensions but also incorporate the intersection of all three DL dimensions into a general DL measure underlying all DLS items, thus suggesting the bifactor structure is likely more aligned with the DL framework on which the DLS instrument is based. In summary, given the unique features of the bifactor model, this alternative structure might be another direction of future research.

5. CONCLUSION

As a summary, the study largely reconfirmed the unidimensional structure of the DLS instrument as was previously reported in the literature (e.g., Ustundag *et al*. (2017)). From the perspective of Rasch measurement theory, the study identified new evidence of reliability and validity to show the DLS instrument is mostly psychometrically sound and therefore is able to produce high quality data measuring digital literacy, which largely supports the findings of the literature that the DLS instrument has a special potential in the research of digital literacy among the Turkish university students. Items demonstrating misfit or DIF were identified which should be further examined and revised using both statistical and nonstatistical criteria through an iterative process.

Acknowledgments

The authors acknowledge the work of the editorial team of the International Journal of Assessment Tools in Education.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Firat University Social Sciences and Humanities Ethics Committee, 23.05.2023-16088.

Contribution of Authors

Hongwei Yang: Conducted the Rasch analysis in the Winsteps software. **Müslim Alanoğlu**: Took responsibility in data collection and related matters. **Songül Karabatak**: Took responsibility in data collection and related matters. **Kelly D. Bradley**: Provided methodological support on Rasch analysis. Finally, all four authors contributed to the writing and finalization of the manuscript.

Orcid

Hongwei Yang https://orcid.org/0000-0003-0225-3792
Müslim Alanoğlu https://orcid.org/0000-0003-1828-4593
Songül Karabatak https://orcid.org/0000-0002-1303-2429
Kelly D. Bradley https://orcid.org/0000-0002-4682-8212

REFERENCES

Adroher, N.D., Prodinger, B., Fellinghauer, C.S., & Tennant, A. (2018) All metrics are equal, but some metrics are more equal than others: A systematic search and review on the use of the term metric'. *PloS ONE, 13*(3), e0193861. https://doi.org/10.1371/journal.pone.0193861

Ala-Mutka, K. (2011). *Mapping digital competence: Towards a conceptual understanding* (Technical Note: JRC 67075). https://doi.org/10.13140/RG.2.2.18046.00322

Amin, H., Malik, M.A., & Akkaya, B. (2022). Development and validation of Digital Literacy Scale (DLS) and its Implication for higher education. *International Journal of Distance Education and E-Learning, 7*(1), 24-43. https://doi.org/10.36261/ijdeel.v7i1.2224

Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory: Measuring in the educational, social and health sciences*. Springer.

Anthonysamy, L., Koo, A.C., & Hew, S.H. (2020). Self-regulated learning strategies in higher education: Fostering digital literacy for sustainable lifelong learning. *Education and Information Technologies, 25,* 2393-2414. https://doi.org/10.1007/s10639-020-10201-8

Anwar, Z., Hanurawan, F., Chusniyah, T., & Setiyowati, N. (2023). Adaptation of the Academic Digital Literacy Scale for college students: A validity and reliability study. *Psychological Science and Education, 28*(4), 98-111. https://doi.org/10.17759/pse.2023280406

Aryadoust, V., Ng, L.Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing, 38*(1), 6-40. https://doi.org/10.1177/0265532220927487

Aydınlar, A., Mavi, A., Kütükçü, E., Kırımlı, E.E., Alış, D., Akın, A., & Altıntaş, L. (2024). Awareness and level of digital literacy among students receiving health-based education. *BMC Medical Education,* 24-38. https://doi.org/10.1186/s12909-024-05025-w

Bartholomew, D.J., Steele, F., Moustaki, I., & Galbraith, J. (2008). *Analysis of multivariate social science data* (2nd ed.). Routledge.

Bond, T.G., & Fox, C.M. (2015). *Applying the Rasch model* (3rd ed.). Routledge.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge University Press.

Cameron, I.M., Scott, N.W., Adler, M., & Reid, I.C. (2014). A comparison of three methods of assessing Differential Item Functioning (DIF) in the hospital anxiety depression scale: Ordinal logistic regression, Rasch analysis and the Mantel chi-square procedure. *Quality of Life Research, 23*(10), 2883-2888. https://doi.org/10.1007/s11136-014-0719-3

Cape, J., Whittington, C., Buszewicz, M., & Underwood, L. (2010). Brief psychological therapies for anxiety and depression in primary care: Meta-analysis and meta-regression. *BMC Medicine, 8,* 38. https://doi.org/10.1186/1741-7015-8-38

Capinding, A.T. (2024). Development and validation of instruments for assessing the impact of artificial intelligence on students in higher education. *International Journal of Educational Methodology, 10*(2), 997 – 1011. https://doi.org/10.12973/ijem.10.2.997

Ceylan, G., Eken, M.O., Yuruk, S., & Emir, F. (2023). Examining the influence of self-esteem and digital literacy on professional competence factors in dental education: A cross-sectional study. *Applied Sciences, 13*(16), 9411. https://doi.org/10.3390/app13169411

Chandra, S., Ghadei, K., Chennamadhava, M., & Ali, W. (2024). Development and validation of a farmer's focused Digital Literacy Scale. *Indian Journal of Extension Education, 60*(1), 111-115. https://doi.org/10.48165/IJEE.2024.601RT1

Chen, F.F., Hayes, A., Carver, C.S., Laurenceau, J.P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality, 80*(1), 219–251. https://doi.org/10.1111/j.1467-6494.2011.00739.x

Da Dalt, L., Anselmi, P., Bressan, S., Carraro, S., Baraldi, E., Robusto, E. & Perilongo, G. (2013). A short questionnaire to assess pediatric resident's competencies: the validation process. *Italian Journal of Pediatrics, 39*(41), 1-8. https://doi.org/10.1186/1824-7288-39-41

Da Dalt, L., Anselmi, P., Furlan, S., Carraro, S., Baraldi, E., Robusto, E., & Perilongo, G. (2015). Validating a set of tools designed to assess the perceived quality of training of pediatric residency programs. *Italian Journal of Pediatrics, 41*(2), 1-7. https://doi.org/10.1186/s13052-014-0106-2

Durak, H.Y., & Seferoğlu, S.S. (2020). Antecedents of social media usage status: Examination of predictiveness of digital literacy, academic performance, and fear of missing out variables. *Social Science Quarterly, 101*(3), 1056-1074. https://doi.org/10.1111/ssqu.12790

Erol, S., & Aydin, E. (2021). Digital literacy status of Turkish teachers. *International Online Journal of Educational Sciences, 13*(2), 620-633. https://doi.org/10.15345/iojes.2021.02.020

Fan, J., & Bond T. (2019). Applying Rasch measurement in language assessment: Unidimensionality and local independence. In V. Aryadoust & M. Raquel (Eds.), *Quantitative data analysis for language assessment*, Vol. I: Fundamental techniques (pp. 83–102). Routledge. https://doi.org/10.4324/9781315187815

Frampton, G.K., & Shepherd, J. (2011). Patient-reported outcomes in clinical trials of inhaled asthma medications: Systematic review and research needs. *Quality of Life Research, 20,* 343–357. https://doi.org/10.1007/s11136-010-9750-1

Garzon, J., & Garzon, J. (2023). Teachers' digital literacy and self-efficacy in blended learning. *International Journal of Multidisciplinary Educational Research and Innovation. 1*(4), 162-174. https://doi.org/10.17613/cmjv-1386

Gignac, G.E., & Kretzschmar, A. (2017). Evaluating dimensional distinctness with correlated-factor models: Limitations and suggestions. *Intelligence, 62,* 138-147. https://doi.org/10.1016/j.intell.2017.04.001

Gillen, J., & Barton, D. (2010). *Digital literacies: A research briefing by the technology enhanced learning phase of the Teaching and Learning Research Programme*. London Knowledge Lab, Institute of Education.

Gilster, P. (1997). *Digital literacy*. John Wiley & Sons.

Gocen, A., & Sen, S. (2021). A validation of Servant Leadership Scale on multinational sample. *Psychological Reports, 124*(2), 752-770. https://doi.org/10.1177/0033294120957246

Gutiérrez-Ángel, N., Sánchez-García, J.-N., Mercader-Rubio, I., García-Martín, J., & Brito-Costa, S. (2022). Digital literacy in the university setting: A literature review of empirical studies between 2010 and 2021. *Frontiers in Psychology, 13,* 896800. https://doi.org/10.3389/fpsyg.2022.896800

Hagquist, C. (2019). Explaining differential item functioning focusing on the crucial role of external information–an example from the measurement of adolescent mental health. *BMC Medical Research Methodology, 19*(185), 1-9. https://doi.org/10.1186/s12874-019-0828-3

Hagquist, C., Bruce, M., & Gustavsson, J.P. (2009). Using the Rasch model in nursing research: An introduction and illustrative example. *International Journal of Nursing Studies, 46*(3), 380-393. https://doi.org/10.1016/j.ijnurstu.2008.10.007

Hamilton, C.B., & Chesworth, B.M. (2013). A Rasch-validated version of the upper extremity functional index for interval-level measurement of upper extremity function. *Physical Therapy, 93*(11), 1507–1519. https://doi.org/10.2522/ptj.20130041

Hamutoğlu, N.B., Canan Güngören, Ö., Kaya Uyanık, G., & Gür Erdoğan, D. (2017). Adapting digital literacy scale into Turkish. *Ege Journal of Education, 18*(1), 408–29. https://dergipark.org.tr/tr/download/article-file/326304

Horn, J.L., & McArdle, J.J. (1992). A practical and theoretical to measurement invariance in Aging research. *Experimental Aging Research, 18,* 117-144. https://doi.org/10.1080/03610739208253916

Hwang, H.S., Zhu, L.C., & Cui, Q. (2023). Development and validation of a Digital Literacy Scale in the artificial intelligence era for college students. *KSII Transactions on Internet and Information Systems, 17*(8), 2241-2258. https://doi.org/10.3837/tiis.2023.08.016

Lazonder, A.W., Walraven, A., Gijlers, H., & Janssen, N. (2020). Longitudinal assessment of digital literacy in children: Findings from a large Dutch single-school study. *Computers & Education, 143,* 103681. https://doi.org/10.1016/j.compedu.2019.103681

Lee, Y.W. (2004). Examining passage-related local item dependence (LID) and measurement construct using Q3 statistics in an EFL reading comprehension test. *Language Testing, 21*(1), 74–100. https://doi.org/10.1191/0265532204lt260oa

Linacre, J.M. (n.d.). *Table 23.0 Variance components for items*. https://www.winsteps.com/winman/table23_0.htm

Linacre, J.M. (2018, September 2). *Detecting multidimensionality in Rasch data using Winsteps Table 23* [Video]. Youtube. https://www.youtube.com/watch?v=sna19QemE50

Linacre, J.M. (2023). *Winsteps® Rasch measurement computer program user's guide*. Version 5.6.4. Winsteps.com

Liu, Y., Millsap, R.E., West, S.G., Tein, J.-Y., Tanaka, R., & Grimm, K.J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods, 22*(3), 486–506. https://doi.org/10.1037/met0000075

Liza, K., & Andriyanti, E. (2020). Digital literacy scale of English pre-service teachers and their perceived readiness toward the application of digital technologies. *Journal of Education and Learning, 14*(1), 74-79. https://doi.org/10.11591/edulearn.v14i1.13925

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invaria nce. *Psychometrika, 58,* 525-543. https://doi.org/10.1007/BF02294825

Miot, H.A. (2020). Analysis of ordinal data in clinical and experimental studies. *Jornal Vascular Brasileiro, 19,* e20200185. https://doi.org/10.1590/1677-5449.200185

Mundfrom, D.J., Shaw, D.G., & Ke, T.L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing, 5*(2), 159–168. https://doi.org/10.1207/s15327574ijt0502_4

Ng, W. (2012a). Can we teach digital natives digital literacy?. *Computers & Education, 59*(3), 6881065-1078. https://doi.org/10.1016/j.compedu.2012.04.016

Ng, W. (2012b). *Empowering scientific literacy through digital literacy and multiliteracies*. Nova Science Publishers, Inc.

Nguyen, L.A.T., & Habók, A. (2024). Tools for assessing teacher digital literacy: A review. *Journals of Computers in Education, 11,* 305-346). https://doi.org/10.1007/s40692-022-00257-5

Ningsih, S.K., Mulyono, H., Rahmah, R.A., & Fitriani, N.A. (2021). A Rasch-based validation of EFL teachers' received online social support scale. *Cogent Education, 8*(1), 1-13. https://doi.org/10.1080/2331186X.2021.1957529

Noorrizki, R.D., Abadi, D., Siwi, N.S.W., Sa'id, M., Mantara, A.Y., & Ramadhani, F. (2022). Factors affecting digital literacy in young adults. *Proceedings of the International Conference of Psychology 2022 (ICoPsy 2022), 308-315,* KnE Social Sciences. https://doi.org/10.18502/kss.v7i18.12396

Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Olur, B., & Ocak, G. (2021). Digital literacy self-efficacy scale: A scale development study. *African Educational Research Journal, 9*(2), 581-590. https://doi.org/10.30918/AERJ.92.21.074

Pangrazio, L., Godhe, A.-L., & Ledesma, A.G.L. (2020). What is digital literacy? A comparative review of publications across three language contexts. *E-Learning and Digital Media, 17*(6), 442-459. https://doi.org/10.1177/2042753020946291

Perera, M.U., Gardner, L., & Peiris, A. (2016). *Investigating the Interrelationship between undergraduates' digital literacy and self-regulated learning skills*. Proceedings of the 2016 International Conference on Information Systems (ICIS), Dublin, Ireland.

Reddy, P., Chaudhary, K., Sharma, B., & Chand, D. (2021). Contextualized game-based intervention for digital literacy for the Pacific Islands. *Education and Information Technologies, 26,* 5535–5562. https://doi.org/10.1007/s10639-021-10534-y

Reddy, P., Chaudhary, K., Sharma, B., & Hussein, S. (2023). Essaying the design, development and validation processes of a new digital literacy scale. *Online Information Review, 47*(2), 371-397. https://doi.org/10.1108/oir-10-2021-0532

Reddy, P., Chaudhary, K., & Hussein, S. (2023). A digital literacy model to narrow the digital literacy skills gap. *Heliyon, 9,* e14878. https://doi.org/10.1016/j.heliyon.2023.e14878

Reise, S.P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S.P., Morizot, J., & Hays, R.D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*(1), 19–31. https://doi.org/10.1007/s11136-007-9183-7

Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150. https://doi.org/10.1037/met0000045

Setiyowati N., & Razak, A.Z.A. (2020). Perceived leadership styles and academicians' job performance: Teaching, research, and community services in Indonesia. *Malaysia Online Journal of Psychology and Counseling, 7*(1), 11-26. https://jupidi.um.edu.my/index.php/MOJC/article/view/24179/11654

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Chapman Hall/CRC.

Tor, D.L., Başaran, S.D., & Arık, E. (2022). Examining the digital Literacy level of teacher candidates. *Journal of Kırşehir Education Faculty, 23*(2), 2027-2064. https://doi.org/10.29299/kefad.1047590

UNESCO. (2018). *A global framework of reference on digital literacy skills for indicator 4.4.2*. https://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf

Urbancikova, N., Manakova, N., & Ganna, B. (2017). Socio-economic and regional factors of digital literacy related to prosperity. *Quality Innovation Prosperity, 21*(2), 124-141. https://doi.org/10.12776/qip.v21i2.942

Üstündağ, M.T., Güneş, E., & Bahçivan, E. (2017). Turkish adaptation of Digital Literacy Scale and investigating pre-service science teachers' digital literacy. *Journal of Education and Future, 12,* 19-29. https://dergipark.org.tr/tr/download/article-file/332115

Wright, B.D. (1996). Local dependency, correlations and principal components. *Rasch Measurement Transactions, 10*(3), 509–511. https://www.rasch.org/rmt/rmt103b.htm

Wu, M., DeWitt1, D., Alias, N., & Nazry, N.N.M. (2022). *Digital literacy from 2018-2021: A scientometric study of the literature*. https://doi.org/10.21203/rs.3.rs-1848300/v1

Yen, W.M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8,* 125-145. https://doi.org/10.1177/014662168400800201

Zwick, R., Thayer, D.T., & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement, 36*(1), 1-28. https://doi.org/10.1111/j.1745-3984.1999.tb00543.x

Zwick, R. (2012). A review of ETS Differential Item Functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement. *ETS Research Report Series*, i-30. https://doi.org/10.1002/j.2333-8504.2012.tb02290.x