Research Paper

Journal of Language and Translation
Volume 15, Number 1, 2025, (pp.177-188)

Psychometric Properties of the Willingness to Read Questionnaire in an EFL Context

Saeed Mojarradi[1], Masoud Zoghi[2*], Nader Assadi[3]
[1]Ph.D. Candidate, Department of English, Ahar Branch, Islamic Azad University, Ahar, Iran
[2*] Assistant Professor, Department of English, Ahar Branch, Islamic Azad University, Ahar, Iran
[3] Assistant Professor, Department of English, Ahar Branch, Islamic Azad University, Ahar, Iran

Received: January 01, 2024 Accepted: March 28, 2024

Abstract
Willingness to read (WTR) is often assumed to be part of a blanket term known as willingness to communicate (WTC). Although the linguistic bases underlying WTR might be similar to WTC, the theoretical explanations of WTR are most likely to be different from WTC. Given the importance of this construct, the design of an independent, valid, and reliable instrument for WTR measurement seems to be a major requirement in the literature. Using a correlational study, we developed a willingness to read questionnaire (WTRQ) and checked its psychometric properties to ensure the WTRQ's accuracy and appropriateness in an EFL context. In this correlational study, we utilized convenience sampling to recruit our research sample, which comprised 269 participants consisting of EFL learners with varying levels of proficiency. Results obtained from exploratory and confirmatory factor analysis revealed a 5-factor WTRQ with 40 items. Findings also showed that the WTRQ enjoys acceptable psychometric properties in terms of reliability and validity. The study concludes that the WTRQ has the potential to be employed in EFL reading research as a validated instrument for measuring WTR.

Keywords: EFL reading, psychometric properties, willingness to communicate, willingness to read

INTRODUCTION

Research in emotional aspects of language learning has burgeoned over recent years. The field of EFL has now recognized that different aspects of language learning are influenced by a web of feelings and emotions always involved in second or foreign language use (Richards, 2022). A wide spectrum of emotions, such as anxiety, hope, enjoyment, and shame (just to name a few) is currently receiving unprecedented attention in language education. However, EFL learners' willingness to do reading tasks is one of the affective variables that remains under-researched. Relevant research has consistently recognized the importance of motivation and willingness in reading development (Cox et al., 2004; Gambrell, 2011; Wigfield et al., 2004), emphasizing the fact that if students are unwilling to read, they will consider reading activities of little value (Guthrie et al., 2004). Unwilling learners are usually disengaged from reading, and this consequently leads to lower reading comprehension (Moomaw, 2013).

As the evidence shows, when Iranian high school students pursue their studies at an academic level, they are required to read technical/specialized texts in English, which they find challenging and the majority of them are unwilling to read those materials (Khajavy & Ghonsooly, 2017). According to Kirchner & Mostert (2017), learners who are willing and able to engage in literacy-based activities (especially reading), have an advantage over less skillful and reluctant readers. As such, EFL

*****Corresponding Author's Email:
masoud.zoghi@iau.ac.ir

reading researchers have focused on reading motivation and its various manifestations such as WTR (Grabe & Stoller, 2019). Although there are various reasons for EFL learners' poor performance in reading, willingness to read foreign language texts plays an important part in the classroom learning context (Grabe & Stoller, 2019). Past studies show that reluctant EFL readers may not comprehend fully the material taught in the classroom (Moomaw, 2013) and thus may be less engaged (Atef-Vahid & Fard Kashani, 2011). We can address the issue of learners' unwillingness in EFL reading if we know the factors affecting WTR; we need to gain an understanding of the nature of WTR in EFL. Among the different ways to enhance WTR, we suggest developing an instrument that can properly capture this construct.

The present study regarded WTR as a stand-alone construct that could have its specific constituents as found in the recent grounded-theory research of Mojarradi et al. (in press). They conceptualized WTR in EFL as a multidimensional construct comprising five interrelated themes, i.e., bottom-up drives, top-down drives, step-down forces, classroom-related drives, and wanting to read. Based on their findings, text type, text difficulty, and text familiarity are constituents of bottom-up drives which can be seen as learner-external factors affecting EFL reading. Top-down drives are learner-specific strategies; for instance, EFL learners' resort to reading strategies and peer reading to solve the problems they encounter in different situations. WTR arises in response to these strategies. Step-down forces are considered intervening conditions, which have a diminishing influence on reading interest. Quality of textbooks, scoring criteria, and time limitations are among the step-down forces that can reduce reading interest and motivation. Classroom-related drives, such as teacher and classroom impact, communication opportunities, and teaching methods, are contextually causal conditions affecting EFL reading. Wanting to read is the last component of WTR which highlights learners' tendency and interest in reading. It is the outcome or rather the consequence of the actions or responses provided by the EFL learner.

After several runs of literature review, it became evident that except for Mojarradi et al.'s work, WTR has not been treated as a separate concept in the EFL context and thereby an appropriate instrument does not exist for its measurement. We do believe that the field of EFL reading needs a WTR instrument that is psychometrically sound. As it is the first instrument to measure WTR in an EFL context, the merit of a WTR instrument lies in its educational value for EFL reading educators and researchers; it can help them know what factors could affect WTR and act accordingly. Early identification of learners who are not willing to read in EFL could make it possible for reading educators to take appropriate action. The instrument could also indicate that specific areas of EFL reading instruction require change and/or remedial action. Put differently, the results obtained from the instrument may guide practitioners toward more effective delivery of the curriculum.

Last but not least, the WTR instrument can help WTR researchers overcome the issue of the "valid-test" fallacy; it occurs when "researchers adopt pre-existing measures wholesale, simply because they seemed valid in other studies or measurement contexts" (Norris & Ortega, 2003, p. 551). As an example, a few lines of research studies have addressed WTR in the Iranian EFL contexts and its potential association with reading (Borsipour et al., 2019; Khajavy & Ghonsooly, 2017), using the only existing four-part questionnaire developed by MacIntyre et al. (2001). This instrument was originally constructed to measure WTC with a 6-item subscale for assessing foreign language learners' willingness in reading tasks in French. Moreover, the questionnaire's conceptual and empirical construct validity have not explicitly been confirmed. While acknowledging the researchers' contribution, it is evident that the said instrument cannot capture various dimensions of WTR, particularly in EFL. Given the psychological and educational measurement literature (Norris & Ortega, 2003), the conceptualization of the intended construct needs to correspond to the measuring instrument if we are to obtain trustworthy findings and avoid mismeasuring the construct.

The factors involved in WTR (i.e., bottom-up drives, top-down drives, step-down forces, classroom-related drives, and wanting to read) have been identified in a qualitative grounded-theory study. As this finding needs statistical support, it was deemed imperative to carry out a factor analytic study to assess and establish the construct validity of a WTR questionnaire (WTRQ) for use among Iranian EFL learners. Thus, the main purpose of this research was to extend the existing WTR research in an EFL context by testing the psychometric properties and the factor structure of the WTRQ using two techniques commonly employed in scale development – exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). We, therefore, addressed the following research questions: Is the WTRQ a psychometrically sound instrument for measuring learners' willingness to read in EFL contexts? To what extent is the WTRQ valid and reliable?

METHOD

Research Design

This study was based on correlational research as it involved collecting data to determine whether a relation existed between the latent variable (WTR) and the observed variables (the items on the WTRQ).

Participants

The study's participants, who were studying in three different language institutes at the time of conducting the current research, were male and female EFL learners of different levels of proficiency. Their ages ranged from 14 to 50, and their native language was Azerbaijani Turkish. We followed Mundfrom et al.'s (2005) recommendations for sample size determination in structural equation modeling (SEM). According to these suggestions, a minimum sample size of 200 is required to guarantee an outstanding level criterion for 40 observable variables and five factors, and a minimal sample size of 200 is required to obtain a good level criterion. Given these guidelines, we recruited the participants using the convenience sampling method. There were 269 participants included in the research sample, who underwent both exploratory and confirmatory analysis. We performed a set of statistical comparisons on participants' demographic characteristics to discover whether the samples were equal. No statistical differences between the samples were seen.

Instrumentation

The WTRQ consists of five sections that match the five themes obtained from the qualitative data analysis in Mojarradi et al.'s (in press) research. The five sections are (i) Classroom-related Drives, (ii) Step-down Forces, (iii) Bottom-up Drives, (iv) Top-down Drives, and (v) Wanting to Read. To be more specific, the initial 5-factor WTRQ included 64 items with a set of Likert-scale responses, ranging from 'strongly agree' to 'strongly disagree'. The range of the scale is 1 to 4, i.e., 4 = strongly agree, 3 = agree, 2 = disagree and 1 = strongly disagree. Scores range from a low of 64 to a high of 256, with higher scores reflecting greater perceived WTR. The WTRQ also included a section for instructions and respondents' demographic information (Appendix A).

Procedure

After the item generation phase, we decided to calculate both the content validity ratio (CVR) and critical validity index (CVI) of the initial draft; they could help ensure the content validity of the WTRQ. To this end, a panel of five EFL reading experts rated the essentialness of each WTRQ item with a Lawshe rating scale (i.e., Essential; Useful, but not Essential; or Not Necessary) (Lawshe, 1975). Values of CVR range from -1 to 1 and the numeric value of 0.50 is the minimum accepted threshold. In contrast to CVR, CVI is the estimated content validity of the whole instrument. CVI is, in fact, the mean CVR of all final items of the instrument. An acceptable CVI for a measure is above 0.70 (Tilden et al., 1990). Additionally, the initial pool of items was piloted on a group of respondents (21 EFL learners) similar to the study's sample to check the comprehensibility of the instrument items. Out of 64 questionnaire items, 21 problematic items (i.e., CVR< 0.50 and CVI < 0.70) were removed. A few changes were also made in terms of diction and wording. As consent for conducting the study was secured

from the participating EFL learners, the WTRQ was distributed among the respondents through an electronic platform known as ePoll (www.epoll.omid.ca). Within a period of four weeks in the academic year of 2020-2021, we administered the initial questionnaires to the participants.

Statistical Data Analysis

Once all questionnaires were completed and returned online, we coded and entered the data into the statistical packages (SPSS 26 and Smart PLS 4.0, respectively) for relevant analysis. Following the common procedures in the testing of factor structure, we obtained an overview of both sample data sets using descriptive statistics. Normality and outliers were checked using mean, standard deviation, skewness, and kurtosis. As there was no strong violation of the assumptions, we considered that the assumptions were fulfilled.

To statistically discover the factor structure of the WTRQ, we conducted EFA and CFA on the scale. Due to its exploratory nature, we first performed a principal component analysis (PCA); PCA was used to verify the instrument's dimensionality. Then, CFA was conducted using Smart PLS. As its name suggests, CFA is a confirmatory statistical technique that is based on structural equation modeling. This analysis helped us to ensure the existence of a previously verified structure. Using the above statistical procedures, we were also able to estimate the internal consistency or reliability and construct validity in terms of convergent validity and divergent validity of the WTRQ.

Model Fit Evaluation

The reliability and validity of the WTRQ were assessed through CFA using PLS-SEM (i.e., partial least squares structural equation modeling method). The reliability estimation of the instrument was evaluated by Cronbach's alpha, composite reliability (CR), and outer loadings. Meanwhile, its convergent validity was evaluated through the average variance extracted (AVE) values and factor loadings. In addition, discriminant validity was determined through the Fornell-Larcker criterion and the hetero-trait-mono-trait ratio of correlations (HTMT).

RESULTS

Exploratory Factor Analysis

The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was also examined. According to Kaiser's (1974) recommendations in which values between 0.8 and 0.9 are great, the KMO analysis verified that the distribution of the values in the sample was adequate for the analysis (KMO=0.892). Further, Bartlett's test of sphericity for the study's data reached statistical significance ($\chi 2 = 18021.549$, $n = 269$, $p < .001$), indicating the factorability of the correlation matrix.

To simplify the interpretation of the factors in the WTRQ, we used an orthogonal rotation method known as varimax rotation. The Kaiser−Guttmann criterion (eigenvalues > 1) demonstrated that five factors (components) accounted for 87.10% of the total variance with Component 1 contributing 32.05% of the variance, Component 2 contributing 20.11%, Component 3 contributing 17.68%, Component 4 contributing 13.42%, and Component 5 contributing 3.85%. The interpretation of the five components was consistent with previous research (Mojarradi et. al., in press). In other words, the components represented classroom-related drives (12 items), step-down forces (9 items), bottom-up drives (10 items), top-down drives (7 items), and wanting to read (2 items). The minimum number of items recommended for a factor in a multidimensional scale is three (Hair et al., 2022); however, two items can also be acceptable for a subscale (Kline, 2016; Yong & Pearce, 2013).

Table 1

Total Variance Explained by Extracted Factors

Component	Eigen values of the actual data	% of Variance	Cumulative %	Eigen values of the simulative data	Means values of the simu-lative data	Percentile values of the simulative data
1	12.8139	32.0346	32.0346	6.5787	1.3984	1.4384
2	8.0475	20.1187	52.1534	5.6933	1.3551	1.386
3	7.0724	17.6811	69.8345	5.1113	1.3229	1.3504
4	5.3683	13.4208	83.2553	4.0081	1.2949	1.319
5	1.5416	3.8541	87.1094	1.5408	1.269	1.2912

Furthermore, employing parallel analysis enabled the identification of the number of latent factors. Through a comparison of eigenvalues derived from random (simulated) data, it becomes apparent that the eigenvalue obtained from the actual dataset surpasses the simulated eigenvalue in the case of 5 factors. Therefore, in line with parallel analysis, the presence of 5 factors is confirmed. Additionally, to double-check the number of the final factors, Cattells' scree test was employed. Cattells' scree test showed a five-factor solution followed by an obvious break in the eigenvalues function.

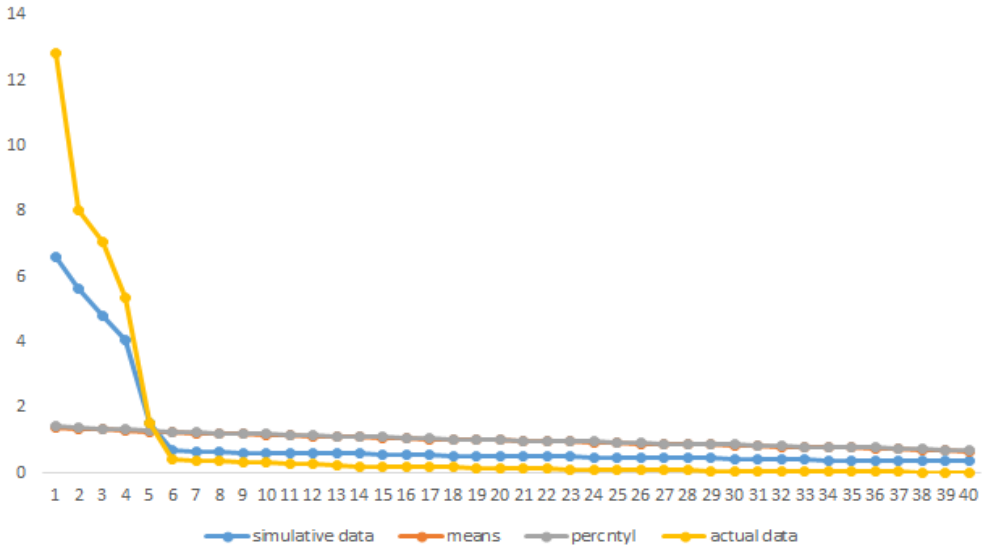

Figure 1

Screen Plot

Based on Hair et. al.'s (2022) recommendations, factor loadings should have at least a loading estimate of 0.5 and ideally exceed 0.7. Therefore, factor loadings of 0.5 or greater were considered acceptable for this research. Items with low communalities do not correlate with other items in the data set. Three items, which did not meet this criterion, were excluded. The final model consisted of 40 items with communalities between 0. 74 to 0.92. The means, standard deviations, and communalities (h^2) for each WTRQ item are shown in Table 2.

We also performed an item-total correlation test to examine to what extent a WTRQ item correlated with the total score of the measure. Results showed that correlation values were within an acceptable range of .30 and .70 (de Vaus, 2004); hence, no item was discarded. Further, in addition to the reliability analysis for each of the five WTRQ subscales, the stratified alpha as an index for internal consistency of the global WTRQ was also calculated. Stratified alpha is intended for scales consisting of stratified layers of items, each of which is assumed to be unidimensional. Rae (2007) recommends that "practicing researchers should routinely calculate both alpha (α) and stratified alpha coefficients (*STRAT*α)". The alphas for each of the components in the measure ($\alpha_{\text{Classroom-related Drives}}$ = 0.986, $\alpha_{\text{Step-down Forces}}$ = 0.983, $\alpha_{\text{Bottom-up Drives}}$ =

0.981, $\alpha_{\text{Top-down Drives}} = 0.977$, and $\alpha_{\text{Wanting to Read}} = 0.819$) were all above the minimum level of .70, recommended by Nunnally (1978). The stratified alpha for the total item pool was found to be acceptable (*STRAT*α= .941). The overall results demonstrate that the WTRQ has decent reliability (Table 3).

Table 2
Summary of EFA Results for WTRQ (n = 269)

Item No.	F1	F2	F3	F4	F5	h^2	*M*	*SD*
q1	**.949**	.052	.112	.090	.008	.923	3.190	0.627
q2	**.921**	.033	.074	.085	-.013	.862	3.164	0.637
q3	**.904**	.065	.072	.090	.047	.837	3.160	0.664
q4	**.911**	-.015	.114	.048	.040	.847	3.171	0.629
q5	**.925**	.067	.104	.113	.021	.884	3.197	0.631
q6	**.947**	.052	.113	.102	.019	.923	3.178	0.639
q7	**.926**	.128	.056	.082	.069	.889	3.167	0.645
q8	**.923**	.006	.105	.107	.046	.877	3.171	0.641
q9	**.929**	.025	.095	.094	.031	.882	3.175	0.654
q10	**.952**	.025	.103	.101	.029	.929	3.186	0.631
q11	**.881**	.113	.053	.103	.040	.804	3.149	0.647
q12	**.884**	.089	.047	.055	.009	.794	3.164	0.649
q13	.124	.022	**.925**	-.007	-.024	.872	2.963	0.742
q14	.040	.039	**.939**	-.012	-.057	.888	2.937	0.733
q15	.078	.078	**.946**	-.030	-.018	.908	2.937	0.717
q16	.101	.073	**.947**	-.007	-.016	.913	2.967	0.735
q17	.101	.089	**.930**	.043	.031	.887	2.967	0.714
q18	.143	.108	**.897**	.021	-.020	.838	2.963	0.722
q19	.136	.085	**.904**	.017	-.039	.844	2.970	0.732
q20	.098	.074	**.953**	.038	.052	.928	2.985	0.728
q21	.074	.059	**.931**	.045	.019	.879	2.955	0.762
q22	.160	.090	.025	**.922**	.078	.891	2.915	0.741
q23	.111	.087	-.008	**.932**	.077	.895	2.915	0.770
q24	.096	.088	-.028	**.922**	.035	.870	2.929	0.737
q25	.147	.063	-.014	**.950**	.028	.928	2.937	0.753
q26	.096	.040	.061	**.935**	.051	.891	2.933	0.735
q27	.097	-.017	.018	**.914**	.026	.847	2.896	0.751
q28	.123	.017	.036	**.915**	.027	.855	2.892	0.758
q29	.057	**.956**	.087	.042	.024	.927	3.119	0.697
q30	.131	**.902**	.088	.024	.009	.838	3.089	0.733
q31	.054	**.936**	.043	.012	.050	.883	3.123	0.672
q32	.100	**.908**	.054	.056	.021	.841	3.093	0.688
q33	.049	**.855**	.073	.051	.033	.743	3.022	0.801
q34	.028	**.910**	.036	.041	.076	.837	3.086	0.736
q35	.045	**.940**	.105	.057	.041	.902	3.126	0.685
q36	.043	**.895**	.069	.046	.065	.814	3.152	0.643
q37	.031	**.930**	.055	.059	-.024	.873	3.123	0.688
q38	.023	**.945**	.034	.031	.081	.902	3.164	0.637
q39	.108	.104	-.025	.081	**.907**	.853	3.506	0.656
q40	.092	.157	-.038	.172	**.882**	.843	3.398	0.648
% of variance	32.05	20.11	17.68	13.42	3.85			
Alpha (α)	.986	.981	.983	.977	.819			
Stratified alpha	.941							

Note. F1 = Classroom-related Drives; F2 = Bottom-up Drives; F3 = Step -down Forces; F4 = Top -down Drives; F5 = Wanting to Read. h^2 denotes communality coefficients. The boldfaced values indicate factor loadings with at least a loading estimate greater than 0.5. *M* = Mean; *SD* = Standard deviation.

Confirmatory Factor Analysis

To provide stronger evidence for the dimensionality of the WTRQ, a two-stage approach known as CFA was used; this allowed for the assessment of the measurement model and the structural model. As shown in Table 3, for each construct on the instrument, the composite reliability (CR) measuring the internal consistency of manifest variables was greater than the reasonable threshold .60. Outer loadings determining an item's contribution to its assigned construct ranged from .58 to .90. Observed variables with outer loading 0.7 or greater are assumed to be completely satisfactory (Gotz et al., 2010). Although a loading value of 0.5 is considered acceptable, indicators with a loading value of less than 0.5 are commonly removed from the instrument (Chin, 1998; Hair et al., 2022).

Cross-validated redundancy assesses the predictive validity of the proposed model using PLS. According to Tenenhaus et al. (2005), a cross-validated redundancy more than zero is considered a predictive model. Goodness-of-fit (GoF) as an overall criterion for model fit was also checked. The range of GoF is between 0 and 1. The results, as shown in Table 3, indicated the appropriate fit of the model (cross-validated redundancy > 0; GoF = .457).

We further examined the construct validity of the WTRQ through convergent and discriminant validity. To this end, we first checked the loading of each manifest indicator on their underlying latent variable (Figure 2). All loadings were greater than the recommended level of 0.5 (Hair et al., 2022) with *t*-values above 2.58 ($p < .01$). (Figure 3)

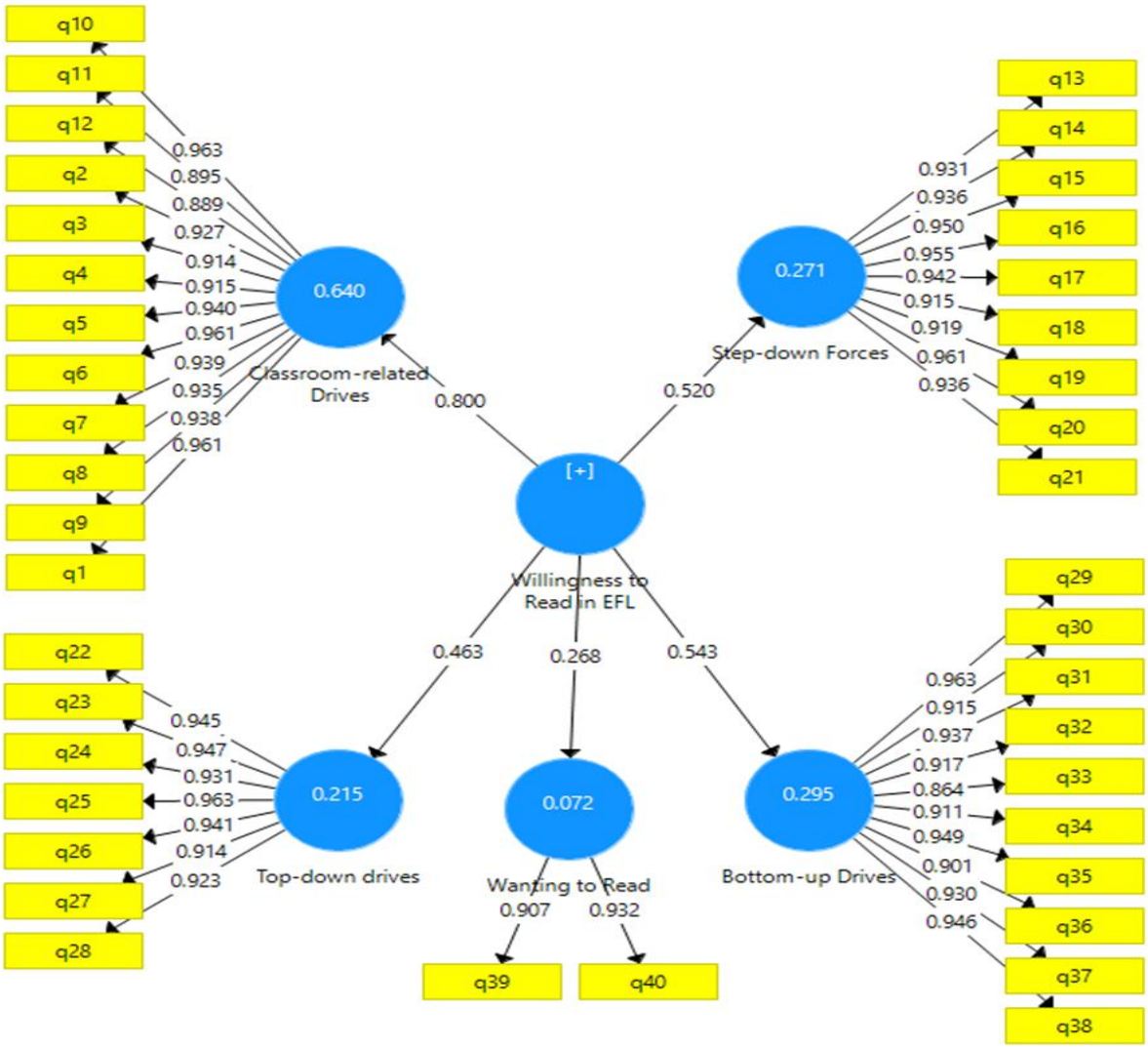

Figure 2

Observed and Latent Variables of the Hypothesized WTRQ model (n = 269)

Figure 3

T-statistics of Factor Loadings of Factors in Confirmatory Factor Analysis

Table 3

CFA Results

Components	Composite Reliability	Average Variance Extracted (AVE)	Cross-validated Communality	Cross-validated Redundancy	GoF
Bottom-up Drives	0.983	0.853	0.811	0.245	
Classroom-related Drives	0.987	0.868	0.837	0.550	
Step-down Forces	0.985	0.881	0.842	0.234	0.457
Top-down drives	0.981	0.879	0.834	0.184	
Wanting to Read	0.917	0.846	0.456	0.058	

According to Fornell and Larcker's (1981) recommendations, researchers can use AVE to determine discriminant validity. Mathematically, it is the square root of the AVE compared with the construct correlations. As Table 4 shows, we found that the AVE of each construct is more than its squared correlation with other constructs. Discriminant validity is also established when the HTMT values are below 0.90 (Henseler et al., 2015). The bootstrapped HTMT results indicated that all the HTMT correlation values ranged between 0.47 and 0.69. As such, adequate support was provided for the discriminant validity of the five-factor WTRQ.

Table 4
Results for Discriminant Validity (Fornell-Larcker Criterion)

	Bottom-up Drives	Classroom-related Drives	Step-down Forces	Top-down drives	Wanting to Read
Bottom-up Drives	0.924				
Classroom-related Drives	0.131	0.932			
Step-down Forces	0.152	0.205	0.939		
Top-down drives	0.113	0.227	0.045	0.938	
Wanting to Read	0.193	0.155	-0.017	0.210	0.920

DISCUSSION

In response to the research question, we developed a 40-item WTRQ with five factors exhibiting strong psychometric characteristics in terms of reliability and validity. Unlike MacIntyre et al.'s (2001) questionnaire, the WTRQ consists of an adequate number of items assessing five factors related to learner-internal and learner-external aspects (the instructor, peers, reading materials, and reading tasks) of WTR. The study's findings offer support for the multidimensionality of WTR reported in Mojarradi et al.'s (in press) research. An interesting point to note is that the current research could demonstrate that statistics is able to confirm and complement the qualitative findings.

Institutions of higher education, language institutes, EFL teachers, and those involved in in-service training can use it to their full potential. They may be introduced to the WTRQ and realize that there are different WTR factors affecting EFL learners' reading performance inside the classroom. While learning about their students' ideas concerning debilitating factors that negatively influence their WTR, they may take practical steps to remove them. In fact, the WTRQ can help EFL teachers assess learners' WTR in a multidimensional way, which is essential for finding an effective way to increase it.

The WTRQ may contribute to a better understanding of WTR and aid EFL teachers in handling unwilling learners in reading tasks inside the classroom. This is especially relevant in EFL contexts like Iran, where a large number of EFL learners are unwilling to read EFL materials. By way of example, if a score increase occurs on the first and second factors of the WTRQ, it may suggest that a change in reading instruction or scoring criteria is necessary; EFL teachers will then realize that they need to adopt an approach with contextual responsiveness to increase learners' reading motivation. As far as the third and fourth components are concerned, an increase in scores on these factors may indicate that teachers need to alter text content or linguistic structures of reading texts to make them more applicable and suitable to students who are learning EFL. It may also suggest that EFL teachers' pay more attention to learner-specific factors, such as reading strategies and a lack of sufficient background or cultural knowledge pertinent to reading material. Additionally, a decrease in the score of the fifth factor may indicate a lack of enjoyment of EFL reading.

Much in the same way, results of the WTRQ can assist EFL reading researchers with their investigations in the field of emotion. For instance, a wide range of studies may be conducted to examine the bivariate or multivariate correlations between WTR and other affective variables such as foreign language enjoyment, reading anxiety, and reading self-efficacy. This could enrich the relevant literature on WTR and contribute to the body of knowledge, particularly in the EFL reading domain.

Despite the use of appropriate methodology for the validation of the WTRQ, there are a few limitations to note. As this is the first study ever conducted to develop an instrument to measure WTR, we could not compare the findings of the study with previous research. It may suggest that further research is warranted as to whether the five-factor model is fit with other data sets in various EFL contexts. This could also provide the opportunity for cross-cultural evaluation of the WTRQ. The cross-culturally validated WTRQ may be safely used in EFL reading research in different EFL contexts. In addition, similar studies can be conducted on WTR outside the classroom to discover to what

extent EFL learners are willing to engage in reading tasks outside the formal learning environment – an issue which deserves reading researchers' attention. The data collection was limited to only three institutes. Therefore, research with samples from other educational institutions such as state-run schools and universities could enhance the generalization of the study's findings. Last but not least, it is worth considering that factor analytic studies are used to determine how well the model fits the data; however, one can argue that a good fit between the assumed model and the observed data does not necessarily mean that the model is correct. A good data-model fit only indicates the plausibility of the model (Schermelleh-Engel et al., 2003). We, therefore, recommend that more research is needed to re-assess the plausibility of the model and the stability of the factor structure underlying the WTRQ.

CONCLUSION

The strong psychometric properties of the WTRQ developed in this study suggest that the instrument enjoys both conceptual stability and psychometric validity and reliability for measuring WTR among EFL learners. With the development of the WTRQ, this research could move the literature forward toward a contextually trustworthy assessment of WTR.

Emotion changes and evolves in various situations for a variety of reasons (Kuppens et al., 2010). In studies on emotion measurement in EFL, attention should therefore be given to the fact that valid instrument construction, for the most part, occurs when an instrument (questionnaire, rating scale, etc.) is developed in a specific context in which it is employed. Situational authenticity often known as contextual validity (Weir, 2005) is used to document and understand the extent to which an instrument yields valid interpretation. The reason is that assessment items generated in a particular context are closer to real-world experiences or tasks described in the descriptors. With the help of previous research (Mojarradi et al., in press), EFL learners' perceptions were transformed into the WTRQ; it is now a promising instrument with situational authenticity to capture the experiences of many learners concerning WTR in an EFL context.

References

Atef-Vahid, S., & Fard Kashani, A. (2011). The Effect of English learning anxiety on Iranian high-school students' English language achievement. *Broad Research in Artificial Intelligence and Neuroscience, 2*(3)29-44. https://lumenpublishing.com/journals/index.php/brain/article/view/1888

Chin, W.W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.). *Modern Methods for Business Research* (pp. 295–336). Erlbaum.

Cox, K. E., Guthrie, J. T., Metsala, J. L., & Wigfield, A. (2004). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading, 3*(3), 231-256.

de Vaus, D. (2004). *Surveys in social research* (5th ed.). Routledge.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.2307/3151312

Gambrell, L. B. (2011). Seven rules of engagement: What's most important to know about motivation to read? *The Reading Teacher, 65*(3), 172-178.

Gotz, O., Liehr-Gobbers, K., & Krafft, M. (2010). Evaluation of structural equation models using the partial least squares (PLS) approach. In V. E. Vinzi, W.W. Chin, J. Henseler, & H. Wang (Eds.). *Handbook of Partial Least Squares* (pp. 691–711). Springer. https://doi.org/10.1007/978-3-540-32827-8_30

Grabe, W., & Stoller, F.L. (2019). *Teaching and researching reading* (3rd ed.). Routledge. https://doi.org/10.4324/9781315726274

Guthrie, J. (2004). Teaching for literacy engagement. *Journal of Literacy Research, 36*(1), 1 – 28.

Hair, J. F., Hult, G. T. M., Ringle, C. M., and Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd Ed.). Sage.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

Kirchner, E. & Mostert, M. L. (2017). Aspects of the reading motivation and reading activity of Namibian primary school readers. *Cogent Education, 4*(1). https://doi.org/10.1080/2331186X.2017.1411036

Khajavy, G.H., & Ghonsooly, B. (2017). Predictors of willingness to read in English: Testing a model based on possible selves and self-confidence. *Journal of Multilingual and Multicultural Development, 1*(2), 1-15. https://doi.org/10.1080/01434632.2017.1284853

Kline, R. B. (2016). *Principles and practise of structural equation modeling* (4th ed.). Guilford.

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology, 99*(6), 1042–1060. https://doi.org/10.1037/a0020962

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563-575.

MacIntyre, P. D., Baker, S. C., Clément, R., & Conrod, S. (2001). Willingness to communicate, social support, and language-learning orientations of immersion students. *Studies In Second Language Acquisition, 23*(3), 369-388. https://doi.org/10.1017/S0272263101003035

Mojarradi, S., Zoghi, M., & Assadi, N. (in press).

Moomaw, J. (2013). *Factors that foster or hinder student reading motivation in a suburban primary school* (M.S. Thesis). http://hdl.handle.net/20.500.12648/5374

Mundfrom, D. J., Shaw, D. G., Lu Ke, T. (2005). Minimum sample size recommendations for conducting factor analysis, *International Journal of Testing, 5*(2), 159-168.

Norris, J., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 717–761). Blackwell. https://doi.org/10.1002/9780470756492

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

Rae, G. (2007). A note on using stratified alpha to estimate the composite reliability of a test composed of interrelated nonhomogeneous items. *Psychological Methods, 12*(2), 177–184. https://doi.org/10.1037/1082-989X.12.2.177

Richards, J. C. (2022). Exploring emotions in language teaching. *RELC Journal, 53*(1), 225–239. https://doi.org/10.1177/0033688220927531

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8*(2), 23–74.

Tenenhaus, M., Vinzi, V. E., Chatelin, Y-M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics and Data Analysis, 48*(1), 159–205.

Tilden, V. P., Nelson, C. A., & May, B. A. (1990). Use of qualitative methods to enhance content validity. *Nursing Research, 39*(3), 172-175.

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave-Macmillan.

Wigfield, A., Guthrie, J. T., Tonks, S., & Perencevich, K. C. (2004). Children's

motivation for reading: Domain specificity and instructional influences. *The Journal of Educational Research, 97*(6), 299-309.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79–94.

Biodata

Saeed Mojarradi received his Bachelor's degree in English Literature from Tabriz National University, followed by his Master's degree in English Language Teaching from Ahar Islamic Azad University. Mojarradi is currently pursuing a doctoral degree at Ahar University. His scholarly contributions include publishing book and articles in esteemed international journals, exploring the effects of utilizing flashcards and English instruction in both public schools and private institutes on students' language proficiency.
Orcid 0000-0001-9149-5720
Email: *saeedmojarradi@outlook.com*

Masoud Zoghi is Associate Professor in the Department of ELT at Ahar Branch, Islamic Azad University, Ahar, Iran. His current research focus includes research methodology, psychology of language, and language assessment.
Orcid 0000-0001-6480-0478
Email: *masoud.zoghi@iau.ac.ir*

Nader Assadi Aidinlou is a lecturer at Islamic Azad University, Ahar Branch in Iran. He has received his B.A. and M.A. in Teaching English as a Foreign Language (TEFL) in Iran and his PhD from University Sains Malaysia. He has been the vice, chancellor for Research and Technology and the chancellor at Islamic Azad University, Ahar Branch for three and four years respectively.
Orcid 0000-0002-2284-8497
Email: *naderasadi@yahoo.com*