Psychology of Language and Communication 2018, Vol. 22, No. 1

DE GRUYTER OPEN

DOI: 10.2478/plc-2018-0002

AMIR H. SARKESHIKIAN, ABDOL-MAJID TABATABAEE,
MARYAM TALEB DOAEE
Department of English Language Teaching, Qom Branch, Islamic Azad University, Qom, Iran

UNIDIMENSIONALITY AND CONSTRUCT VALIDITY OF THE SELF-REGULATING CAPACITY IN VOCABULARY LEARNING (SRCVOC) IN IRANIAN EFL CONTEXT: ITEM-LEVEL RESPONSES VERSUS ITEM PARCELS

This study investigated the psychometric properties of self-regulating capacity in vocabulary learning scale (Tseng, Dornyei, & Schmitt, 2006) in the Iranian EFL context. For this purpose, a sample of 1167 high school students completed the Persian SRCvoc in the main phase. The internal consistency reliability of the scale was examined using Cronbach's alpha. It showed acceptable reliability in both piloting and main phases. The results of exploratory factor analysis (EFA) showed that the SRCvoc is composed of three factors. However, confirmatory factor analysis (CFA) on the three-factor model of the SRCvoc and Tseng et al.'s (2006) five-factor model of the SRCvoc with item-level indicators showed that both models did not fit the data. The findings of this study imply that the item-parcels in Tseng et al. (2006) may have masked the nature of the factor structure of the self-regulating capacity in vocabulary learning scale. It should therefore be re-theorized.

Key words: confirmatory factor analysis, construct validity, exploratory factor analysis, unidimensionality, self-regulating capacity in vocabulary learning (SRCvoc)

Introduction

The concept of academic self-regulation emerged more than two decades ago in educational psychology to answer the question of how students master their learning processes (see Zimmerman, 2008). Over the past decades, second

Address for correspondence: Amir Hossein Sarkeshikian , Department of English Language Teaching, Faculty of Humanities, Pardisan St., Qom, Iran. E-mail: sarkeshikian@qom-iau.ac.ir

language (L2) vocabulary learning has attracted the attention of researchers (e.g., Nation, 2001, 2013). Moreover, research in vocabulary learning strategies gained momentum because most L2 learners have to learn vocabulary independently and outside the classroom in most cases (Mizumoto, 2013). As a result of this realization, vocabulary learning strategies have been extensively studied (e.g., Berns, 2010; Schmitt, 2000; Tseng & Schmitt, 2008). However, despite the bulk of empirical studies on L2 vocabulary learning, the term learning strategy has been conceptualized in different ways (see Cohen, 2014; Oxford, 2017). According to Chamot (2004), learning strategies are conscious thoughts and actions that are used by learners to achieve learning goals. Oxford (2003) stated that learning strategies are one of the elements that may help recognize how learners learn a second or foreign language. In the case of learning L2 vocabulary, it is believed that students should be instructed in how to use vocabulary learning strategies (Schmitt, 2000) because most of them do not take a strategic approach to learn vocabulary (Moir & Nation, 2002).

However, the notion of learning strategies has been incorporated into the broader concept of self-regulated learning (SRL) due to the paucity of firm theoretical foundations (Mizumoto, 2013). To fill this gap in the literature on vocabulary learning strategies, Tseng, Dörnyei, and Schmitt (2006) tried to conceptualize strategic vocabulary learning based on the theories of self-regulation by targeting "the core learner difference that distinguishes self-regulated learners from their peers who do not engage in strategic learning" (p. 80). To that end, they developed and validated a scale "that operationalizes the newly-conceived system of self-regulatory capacity" in vocabulary learning using item-parcels (Tseng, et al., 2006, p. 80). While some researchers have used item-parceling techniques to validate the self-regulating capacity in vocabulary learning scale (SRCvoc) in different contexts (Mizumoto & Takeuchi, 2012; Yesilbursa & Bilican, 2012), few studies have investigated the psychometric properties of Tseng et al.'s (2006) model of the self-regulating capacity in vocabulary learning scale with item-level indicators.

Review of the Related Literature

Until now, several models of SRL have been proposed from different theoretical perspectives. From among them, four models have been considered as the most important ones: the adaptable learning model (Boekaerts & Niemivirta, 2000), general framework for SRL (Pintrich, 2004), four-stage model of SRL (Winne & Hadwin, 1998), and cyclic model of self-regulation (Zimmerman & Camppillo, 2003). Although these models involve slightly different constructs and processes, they all share a basic assumption. The common assumption is that "learners are considered as active participants in their learning potentially monitoring, controlling and evaluating certain aspects of their cognition, behavior, affects and environment for the attainment of their goals" (Hirata, 2010, p. 33).

In line with educational psychology, the field of L2 learning has shifted away from a focus on the teacher and underpinned the role of the learners and their language learning processes (Dörnyei, 2005). Most recently, "the concept of language-learning strategies has expanded into a more extensive notion of self-regulated learning, partly in response to a wave of criticism directed at the paucity of rigid theoretical underpinnings" (Mizumoto, 2013, p. 16). Parallel to this, the focus of language learning research has similarly shifted away from investigating the product of language learning to its processes (Dörnyei & Skehan, 2003).

In the same vein, Tseng et al. (2006) developed and validated the self-regulating capacity in vocabulary learning scale (SRCvoc). The results of item analyses showed that four items did not perform well, and thus they were deleted, leaving 41 items for the subsequent reliability analysis. The results of reliability analysis revealed that the whole questionnaire had acceptable internal consistency reliability (α = 0.78), and each of the SRCvoc subscales showed an alpha coefficient above 0.70. The results of the main phase of their study showed that the reliability indices were only marginally lower than in the pilot sample, with a mean scale coefficient of 0.77. In the third phase, they administered the revised version of the instrument to172 senior high school students from two public schools to check the construct validity of their measure. They used confirmatory factor analysis (CFA) to explore the construct validity of the instrument. The fitness indices showed that "the SRCvoc is a meaningful and valid measure and can serve as a basis for exploring the theoretical nature of self-regulation" (p. 94). Tseng et al. (2006) used exploratory factor analysis (EFA) to examine the unidimensionality of the instrument. The results showed that the SRCvoc measures just one single trait.

Mizumoto and Takeuchi (2012) adapted and validated the SRCvoc (Tseng et al., 2006) in a Japanese EFL setting. They used both EFA and CFA to check the factor structure and construct validity of the scale. For this purpose, they translated the SRCvoc items into Japanese, and then back-translated them into English. They administered the Japanese version of the SRCvoc to 443 EFL learners who were majoring in humanities or engineering at four different universities in western Japan, with an age range of 18-22 (n_{males} = 208, n_{females} = 235). The results of the pilot study showed that two items did not function well. They were deleted, and 18 items were selected to replicate Tseng et al.'s (2006) model. The results of EFA revealed that the factor structure of the scale was different from those in the original study. As a result, 12 items were discarded. The remaining items were administered to 914 EFL learners at five universities in Japan, within the age range of 18-22 (n_{males} = 425, n_{females} = 489). The construct validity of the questionnaire was investigated using CFA. Although the reliability coefficients were rather low as compared with those in the original questionnaire, the results of CFA were all acceptable. They found that the replication of Tseng et al.'s (2006) model

in their study would be unjustifiable, thus they conducted EFA using maximum likelihood with promax rotation to reexamine the factor structure of the SRCvoc. Although the factor structure was different from those suggested in Tseng et al. (2006), this study demonstrated that SRCvoc could be a valid and reliable measure of the volitional aspect of self-regulating capacity in vocabulary learning in a Japanese EFL environment.

In addition, the SRCvoc was validated in Turkey by Yesilbursa and Bilican (2013). The results of this study suggested that the Turkish version of the SRCvoc was a reliable and valid instrument in the Turkish EFL context. They found that the Turkish version of the instrument had high internal consistency reliability (α = .89), in line with the original version in context of Taiwan. To that end, CFA was used to investigate the construct validity of the Turkish version. They ran CFA and found that item 12 (i.e., When I feel stressed about vocabulary learning, I simply want to give up.) and item 15 (i.e., When I feel stressed about my vocabulary learning, I cope with this problem immediately.) had weak consistency, so they deleted these two items. They pointed out that the SRCvoc "may be sensitive to cultural differences, and hence further studies need to be conducted in different cultural contexts with participants of different ages to shed more light on the concept" (p. 885). Presumably, learners' self-regulating capacity is culturally related to their agency, thus it may be affected by the cultural context that learners are embedded in (Trommsdorff , 2009).

As a matter of fact, Tseng et al. (2006) used an item-parceling technique to investigate the psychometric properties of the SRCvoc in terms of dimensionality and construct validity. However, since the use of item-parcels in a CFA model may cause better fitting solutions (Bandalos, 2002), estimation bias (Matsunaga, 2008) and measurement invariance (Meade & Kroustalis, 2006), the current study sought to investigate the five-factor model of the SRCvoc (Tseng et al., 2006) with item-level indicators. Hence, given the culture-dependent nature of self-regulation, the following questions were posed:

RQ1: Is the self-regulating capacity in vocabulary learning (SRCvoc) questionnaire a reliable scale in the Iranian EFL context?

RQ2: Is the self-regulating capacity in vocabulary learning (SRCvoc) questionnaire a valid instrument with five subcomponents in the Iranian EFL context, if item-level indicators are not parceled?

Methods

Participants

The participants of the piloting phase of this study were 43 female students in a high school in Kashan, Iran, where co-educational classes are not held based on the regulations of the Iranian Ministry of Education. This intact group

included sophomores (n = 14), juniors (n = 16) and seniors (n = 13). Their age range was from 15 to 19 (M = 16.32, SD = .92). The participants of the main study were 1167 Iranian high school students (n_{male} = 651, n_{female}= 516). The sample included students from ninth grade (n = 402), tenth grade (n = 260), 11th grade (n = 233), and 12th grade (n = 270). They were from fifteen public schools in three Iranian cities. Their ages ranged from 14 to 20 (M = 16.13, SD = 1.26). In order to increase the transferability of the findings (see Brown, 2006), the age, sex and educational backgrounds of this convenient sample were controlled for.

Instrument

The instrument used in this study was the Persian version of self-regulatory vocabulary learning scale (SRCvoc), developed by Tseng et al. (2006). This scale is a twenty-item questionnaire with five subscales. The subscales of the instrument are as follows: a) commitment control (items 4, 7, 10, and 13) (henceforth CC), b) metacognitive control (items 5, 9, 11, and 16) (henceforth MC), c) satiation control (items 1, 8, 18, and 19) (henceforth SC), d) emotion control (items 2, 6, 12, and 15) (henceforth EC), and e) environment control (items 3, 14, 17, and 20) (henceforth EnC). All the items were based on a six-point Likert scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Respondents were required to tick the appropriate box for the option that best expressed their personal vocabulary learning experience (Tseng et al., 2006). The reason for the selection of this questionnaire was that it was the only available instrument that measures the construct of self-regulation in vocabulary learning.

Procedure

This study followed four procedural steps: (a) translation and adaptation of the SRCvoc scale, (b) piloting the translated version of the questionnaire and designing the final version based on the pilot results, (c) administering the instrument to a large sample of Iranian high school students, (d) conducting the same statistical analyses that were run in Tseng et al. (2006) (i.e., Cronbach alpha, EFA, and CFA).

1. In order to translate the questionnaire into Persian, three steps were taken:
    - Initial translation: The English-Persian translation was performed by the researchers.
    - Back-translation: The initial translation was translated back to English by a Persian-to-English translation expert.
    - Revision and adaptation: After back-translation, another expert made the necessary adjustments to prepare the final Persian questionnaire.

2. After translating the questionnaire to Persian, it was distributed to an intact group of 43 female students in a high school in an Iranian city. The data gathered through this piloting phase showed that the scale had acceptable internal consistency reliability (α = .81). The results of the piloting study showed that item 12 within the emotion control subscale had low item-total value (r_{pbi} <.16), and it was therefore discarded. The revised version was used in the main phase of the study.
3. In the main phase of the study, the researchers were present at each research site to explain the purpose of the study, to make it clear that the results would not have any effect on the students' course grades, and to assure that their personal data would remain confidential. Afterwards, the participants were asked to complete the questionnaire in about 20 minutes. The survey took place in February, 2016. Due to the regulations of the Ministry of Education in the Islamic Republic of Iran, the female researcher was not allowed to be present at a research site where male learners were present. Therefore, five male English teachers were asked to administer the questionnaire to male respondents. From among these teachers, four of them accepted the researcher's request. They administered the scale in ten schools across three Iranian cities.
4. The data collected in the pilot phase and main phase of the study were imported into SPSS 22.0.
5. To examine the internal consistency reliability of the scale, the data sets were subjected to Cronbach's alpha method. Moreover, point-biserial correlation was used for the purpose of item analysis.
6. Finally, the main data set was subjected to EFA and CFA to investigate the dimensionality and construct validity of the SRCvoc.

Results

Reliability Analysis

The Persian version of the SRCvoc showed acceptable internal consistency reliability (α = .81). Moreover, the reliability of each subscale of the questionnaire was examined using Cronbach's alpha method. "Although the suggested magnitude of average item intercorrelations tends to range from .15 to .50 (Briggs & Cheek, 1986; Clark & Watson, 1995) [,] … averages for broad constructs typically fall in the .15 to .30 range" (Schinka & Velicer, 2003, p. 396). Following Hagell, Rosblom, and Palhagen (2001), items with point-biserial correlation coefficients (r_{pbi}) of .16 or higher were accepted as well-functioning items.

The items that reflected the commitment control subscale were items 4 (r_{pbi} =.38), 7 (r_{pbi} =.44), 10 (r_{pbi} = .4), and 13 (r_{pbi} = .49). The commitment control subscale showed acceptable internal consistency reliability (α = 0.65). Items 5 (r_{pbi} = .37), 9 (r_{pbi} = .46), 11 (r_{pbi} = .44), and 16 (r_{pbi} = .48) represented the metacognitive control subscale. The reliability of this subscale examined through Cronbach's alpha was about 0.66.

The items that reflected the satiation control scale were items 1 (r_{pbi} = -.01), 8 (r_{pbi} = .09), 18 (r_{pbi} = .13), and 19 (r_{pbi} = .07). The subscale showed unacceptable internal consistency reliability (α = .42). Moreover, Item 1 displayed a negative corrected item-total correlation value (r_{pbi} = -.01), suggesting a weak relationship with the rest of items. Hence, the item was discarded and the reliability of the subscale was reexamined. Overall, the satiation control subscale showed better internal consistency reliability (α =.69).

Items 2 (r_{pbi} = .51), 6 (r_{pbi} = .5), and 15 (r_{pbi} = .52) of the questionnaire were related to the emotion control subscale. Cronbach's alpha was used to examine the internal consistency of the subscale (α = .69). During piloting it was found that item 12 of the emotion control subscale had a low item-total correlation coefficient (r_{pbi} = .01), and it was subsequently discarded. The items that reflected the environment control scale were items 3 (r_{pbi} = .36), 14 (r_{pbi} = .43), 17 (r_{pbi} = .48), and 20 (r_{pbi} = .45). The reliability of this subscale was examined using Cronbach's alpha (α = .65).

Results of Exploratory Factor Analysis

To answer the second research question, the main data set was subjected to principle axis factoring (PAF) with varimax rotation to probe the underlying structure of the instrument. It should be noted that the assumptions for factor analysis were met as follows: a) adequacy of sample size (KMO = .93 > .60); b) no zero correlations among variables, χ^2 (153) = 6.28, p < .000; c) no perfect correlations among variables. The results of data reduction through PAF with varimax rotation are shown in Table 1.

As shown in Table 1, three factors with eigenvalues larger than 1 were extracted through PAF with varimax rotation, accounting for 50.39 percent of total variance. The results revealed that one factor explained over 36 percent of the total variance, and the eigenvalues of the second and third largest factors were marginal compared to the first one.

As shown in Table 2, the rotated factor loadings revealed that the descriptors in the SRCvoc corresponded with three latent factors with eigenvalues greater than 1. This factorial structure is not consistent with the five-factor model of the SRCvoc, which was theorized by Tseng et al. (2006).

Based on the results in Table 2, the first factor had the largest number of item loadings, including items for commitment control (items 7, 10, and 13), metacognitive control (items 9, 11, and 16), satiation control (items 8, 18, and 19), emotion control (items 2, 6, and 15), and environment control

(item 14). These thirteen items showed a high internal consistency reliability (α = .88). Therefore, all items were retained.

Table 1. Total Variance Explained for SRCvoc

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.51	36.21	36.21	5.93	32.97	32.97	4.20	23.38	23.38
2	1.44	8.00	44.22	0.93	5.17	38.14	1.95	10.86	34.24
3	1.11	6.17	50.39	0.57	3.18	41.32	1.27	7.08	41.32
4	0.86	4.81	55.21						
5	0.83	4.63	59.84						
6	0.77	4.28	64.12						
7	0.70	3.92	68.04						
8	0.67	3.74	71.79						
9	0.65	3.61	75.41						
10	0.60	3.36	78.78						
11	0.57	3.18	81.96						
12	0.55	3.10	85.06						
13	0.51	2.85	87.92						
14	0.50	2.80	90.72						
15	0.45	2.50	93.22						
16	0.42	2.34	95.56						
17	0.42	2.33	97.90						
18	0.37	2.09	100.00						

The second factor was loaded by three items that represented commitment control (item 4), metacognitive control (item 5), and environment control (item 3). The results of reliability analysis showed that these items yielded a weak internal consistency reliability value (α = .65). Finally, items 17 and 20 clustered together as a separate factor (i.e., environment control). Cronbach's alpha value for these two items was equal to the value obtained for the four items of this subscale in the pilot phase (α = .65).

Confirmatory Factor Analysis for the Three-Factor Model

Following Tseng, Liu, and Nix (2017), CFA was run to investigate the goodness-of-fit of the three-factor model based on the factorial structure obtained from the preceding EFA (Figure 1). The maximum likelihood method was adopted to estimate the parameters involved in the three-factor model.

Table 2. Standardized Factor Loadings

Item	Factor		
	1	2	3
item2	.52		
item6	.47		
item7	.44		
item8	.50		
item9	.53		
item10	.38		
item11	.47		
item13	.59		
item14	.54		
item15	.70		
item16	.55		
item18	.66		
item19	.66		
item3		.39	
item4		.64	
item5		.61	
item17			.69
item20			.68

Note. Factor loadings < .3 were suppressed.

Table 3 displays the standardized regression weights, connecting observed variables to latent factors.

As shown in Table 3, all of the items had significant contributions to their underlying factors, with the exception of items for the second factor. The standardized estimates ranged from a high of .87 for item 15 to a low of .01 for item 3. At the same time, items for the second factor had less than the expected minimum value of .30 (i.e., items 3, 4, and 5). Moreover, the global indices for goodness of fit for this model are displayed in Table 4.

The chi-square test revealed the badness-of-fit of the three-factor model, χ^2 (133) = 765.75, p = .000. Since the chi-square test is sensitive to sample size and almost always turns out to be significant, its ratio over the degree of freedom should be reported (Dattalo, 2013). Based on the result of this ratio, this model did not fit the data (χ^2/df=5.75 > 3). Moreover, the root mean square of error approximation (RMSEA) was well within the satisfactory range (0.05 < RMSEA = 06 < 0.8). However, the PCLOSE test indicated that the null hypothesis (i.e., RMSEA value is not greater than .05) cannot be rejected

(p = .000). The global goodness-of-fit indices were all lower than .95, proving the badness-of-fit of this model (CFI = .89, NFI = .88, TLI = .87, and RFI = .84).

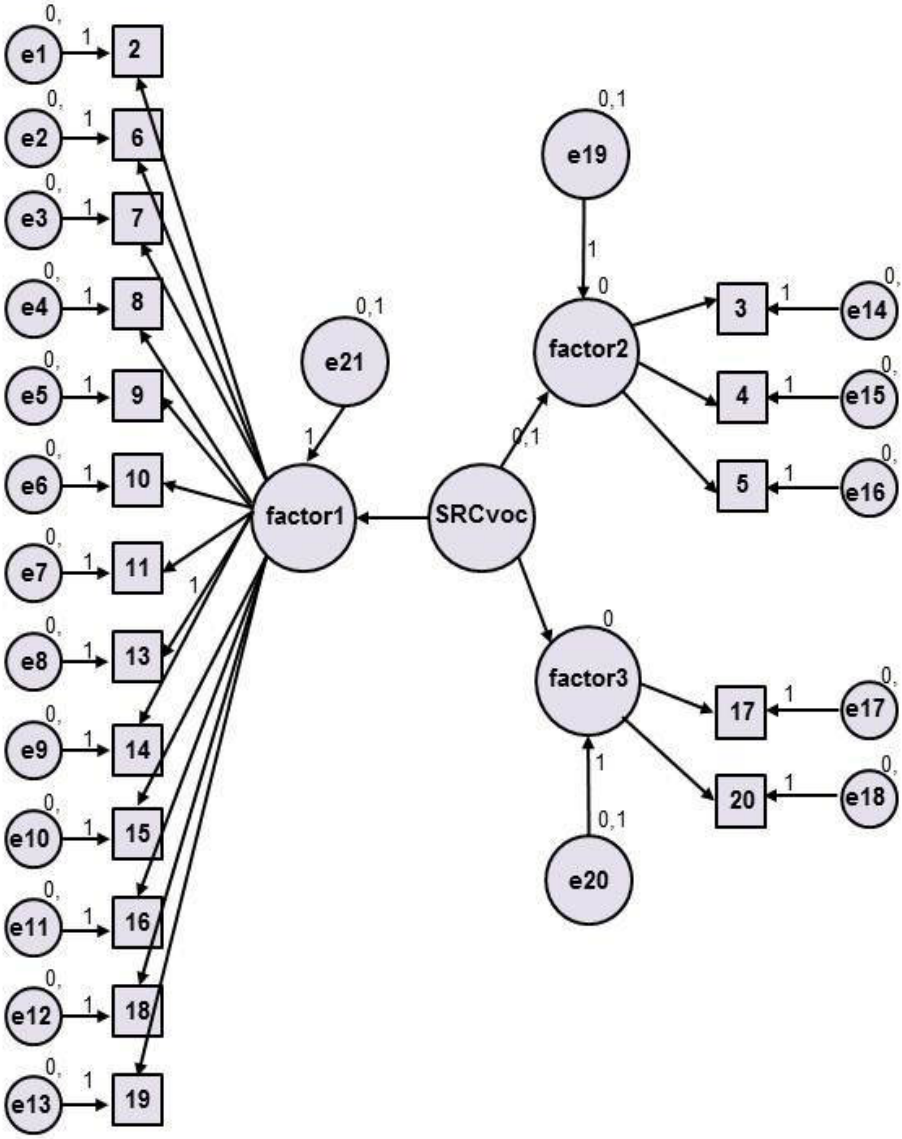

Figure 1. The three-factor model of SRCvoc.

Table 3. Standardized Regression Coefficients

Regression path			Estimate	*p*
fac3	<---	SRCvoc	.54	***
fac1	<---	SRCvoc	.72	***
fac2	<---	SRCvoc	39.59	.99
item13	<---	fac1	1.00	
item11	<---	fac1	.62	***
item10	<---	fac1	.55	***
item9	<---	fac1	.69	***
item17	<---	fac3	.78	***
item20	<---	fac3	.84	***
item3	<---	fac2	.01	.99
item4	<---	fac2	.02	.99
item5	<---	fac2	.02	.99
item8	<---	fac1	.65	***
item7	<---	fac1	.66	***
item6	<---	fac1	.79	***
item2	<---	fac1	.80	***
item14	<---	fac1	.73	***
item15	<---	fac1	.87	***
item16	<---	fac1	.66	***
item18	<---	fac1	.79	***
item19	<---	fac1	.86	***

Note. Item 13 is fixed at 1.00.
***$p < .001$, two-tailed.

Table 4. Fit Indices of the Three-Factor Model of SRCvoc

Variable	X^2	*df*	*p*	$\chi 2/df$	CFI	NFI	TLI	RFI	RMSEA	PCLOSE
Model	765.75	133	.000	5.75	.89	.88	.87	.84	.06	.000

Note. CFI = comparative fit index; NFI = normed fit index; TLI = Tucker-Lewis index; RFI = relative fit index; RMSEA = root mean square error of approximation; PCLOSE = close fit.
***$p < .001$, two-tailed.

Confirmatory Factor Analysis on Factor Structure of Tseng et al.'s (2006) Model

CFA was run to probe the factorial structure of the SRCvoc with five underlying factors, as theorized in Tseng et al. (2006). For the purpose of this study, the items of this model were not parceled, and were hypothesized as the indicators of the model (Figure 2).

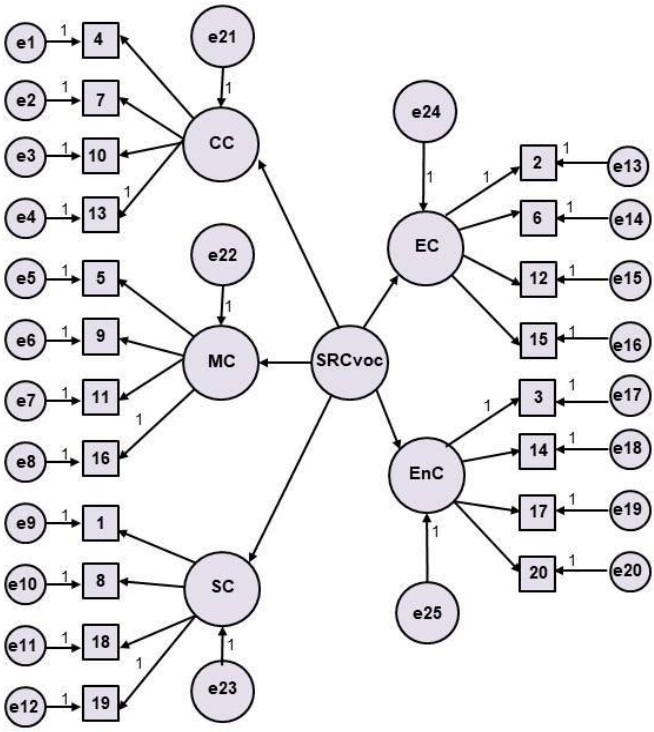

Figure 2. The five-factor model of the SRCvoc (adopted from Tseng et al., 2006).

Table 5 displays the standardized regression weights that connect the observed variables to latent factors. Based on these results, it can be concluded that all of the items had significant contributions to their underlying factors with the exception of items for metacognitive control.

As shown in Table 5, the five factors of the model had significant contributions to the model, with the exception of the MC subscale. The standardized estimates ranged from a high of .51 for item 19 to a low of .08 for item 5. At the same time, only five items had higher values than the minimum excepted value of .30 (i.e., items 14, 15, 18, 19, and 20). Moreover, items 1 and 12 showed negative standardized values, indicating divergence from their associated factors. The indexes for goodness of fit for the first model, displayed in Figure 1, showed that the model did not fit the data (Table 6).

Table 5. Standardized Regression Coefficients

Regression path			Estimate	*p*
CC	<---	SRCvoc	6.73	***
MC	<---	SRCvoc	14.52	.16
SC	<---	SRCvoc	4.49	***
EC	<---	SRCvoc	6.67	***
EnC	<---	SRCvoc	3.36	***
item4	<---	commitment	.19	***
item7	<---	commitment	.22	***
item10	<---	commitment	.18	***
item13	<---	commitment	.27	***
item5	<---	metacognitive	.08	.16
item9	<---	metacognitive	.10	.16
item11	<---	metacognitive	.09	.16
item16	<---	metacognitive	.14	.16
item8	<---	satiation	.31	***
item18	<---	satiation	.40	***
item19	<---	satiation	.51	***
item2	<---	emotion	.27	***
item6	<---	emotion	.27	***
item15	<---	emotion	.34	***
item3	<---	environment	.37	***
item14	<---	environment	.48	***
item17	<---	environment	.28	***
item20	<---	environment	.40	***
item1	<---	satiation	-.10	***
item12	<---	emotion	-.12	***

Note. ***$p < .001$, two-tailed.

As shown in Table 6, the chi-square test was significant, showing the badness-of-fit of the five-factor model, χ^2 (170) = 2931.70, $p = .000$). Moreover, the ratio of the chi-square value over the degree of freedom also showed that this model did not fit the data because it was larger than the recommended value ($\chi^2/df = 17.24 > 3$). Moreover, the root mean square of error approximation (RMSEA) was not within the acceptable range ($0.05 < \text{RMSEA} = .16 > 0.8$). As displayed in Table 6, the results of other goodness-of-fit indices revealed

that Tseng et al.'s (2006) five-factor model with item-level indicators did not fit the data because they were lower than .95 (CFI = .58, NFI = .56, TLI = .48, and RFI=.58).

Table 6. Fit Indices for Tseng et al.'s (2006) Five-Factor Model with Item-Level Indicators

Variable	X^2	*df*	*p*	χ^2/df	CFI	NFI	TLI	RFI	RMSEA
Five-factor model	2931.70	170	.000	17.24	.58	.56	.48	.58	.16

Note. CFI = comparative fit index; NFI = normed fit index; TLI = Tucker-Lewis index; RFI = relative fit index; RMSEA = root mean square error of approximation.
***$p < .001$, two-tailed.

Discussion

This study was aimed at investigating the reliability, unidimensionality and construct validity of the SRCvoc in an Iranian EFL context. In line with the findings of Mizumoto and Takeuchi's (2012), the reliability coefficients of the subscales of this instrument were lower than those reported in Tseng et al. (2006). Furthermore, the item-level analysis showed that item 12 was not a well-functioning item, and was discarded after EFA. Consistent with the results of Mizumoto and Takeuchi (2012) and Yesilbursa and Bilican (2013), this finding raises the need to adapt this instrument.

Moreover, the results of EFA showed that the construct of the SRCvoc is composed of three underlying factors (see Table 2). It is also noteworthy that thirteen items from all five subscales of the SRCvoc loaded on one factor in the three-factor solution, and three items from three different subscales loaded on the second factor, while only two items from environment control loaded on their associated factor. This finding casts doubt on the theoretical framework of the SRCvoc (Tseng et al., 2006), because this factorial structure supports the supposition that self-regulating capacity in vocabulary learning measures a unitary construct. Therefore, the replication of Tseng et al.'s (2006) model is not justifiable because the factorial structure of the SRCvoc in Tseng et al.'s (2006) study is not consistent with the factor structure of the SRCvoc obtained in this study. It is very likely that the item-parceling technique in Tseng et al.'s study altered the nature of the unidimensionality of the SRCvoc, and may have caused estimation bias (Matsunaga, 2008).

Most importantly, the CFA results revealed that the three-factor and five-factor models of SRCvoc did not fit the data. The first item-level model represented the construct of the SRCvoc based on the factor structure obtained in this study (see Figure 1). The results showed that the model did not fit the data (see Table 4). The second model was hypothesized based on the five-factor model of the SRCvoc (see Tseng et al., 2006), but the item indicators of the model were not parceled (see Figure 2). Based on the results, this model did not

converge with satisfactory goodness-of-fit indexes, as compared to those reported in Tseng et al. (2006) (see Table 9). In line with Mizumoto and Takeuchi's (2012) claim, it seems that item parcels in Tseng et al.'s (2006) study must have improved the goodness-of-fit of Tseng et al.'s (2006) model (see Matsunaga, 2008). Overall, the results for both CFA models showed that the psychometric properties of the original SRCvoc (Tseng et al., 2006) must have been masked by item-parceling.

In the literature on psychometrics, it is noted that item parceling can improve model estimation and fit (Matsunaga, 2008). Moreover, the use of parcels as indicators in a CFA model can affect tests of measurement invariance (Meade & Kroustalis, 2006). According to Bandalos (2002), the use of item parcels results in "better fitting solutions, as measured by the root mean squared error of approximation (RMSEA), comparative fit index (CFI), and chi-square test" (p.78). The findings of this study suggest that the SRCvoc suffers from some sources of model misfit due to the use of item parcels in the original model, as hypothesized by Tseng et al. (2006). A possible future trend in investigating the psychometric properties of the SRCvoc may consider the re-theorization of the construct and re-specification of the SRCvoc model. Future research can also test the construct validity of this scale in a larger sample size with more consistent demographics. Moreover, other CFA models of SRCvoc can be hypothesized and run to investigate the effect of such moderating variables as gender, age, educational background, and context.

Conclusions

To conclude, it is necessary to revise the theoretical underpinnings of the self-regulatory capacity in vocabulary learning scale (Tseng et al., 2006). In other words, the model of self-regulated vocabulary learning should be investigated more deeply because this volitional model is part of a complex model of SRL (Zimmerman & Schunk, 2001), and may not be appropriate for the construct of this scale. Moreover, it is concluded that parceling techniques can alter the real psychometric properties of the measurement scales; therefore, it is necessary to consider their detrimental effects on the development and validation of measurement instruments.

References

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*. *9*, 78–102. doi: 10.1207/S15328007SEM0901_5

Berns, M., (2010). *Concise encyclopedia of applied linguistics*. Cambridge: Elsevier.

Boekaerts, M. & Niemivirta, M. (2000). Self-regulated learning: Finding a balance between learning goals and ego-protective goals. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation: Theory, research, and applications* (pp. 417–450). San Diego, CA: Academic Press.

Brown, J. D. (2006). Generalizability from second language research samples. *JALT Testing & Evaluation SIG Newsletter*. *10*(2), 21–24.

Chamot, A. U. (2004). Issues in language learning strategy research and teaching. *Electronic Journal of Foreign Language Teaching*. *1*(1), 14–26.

Cohen, A. D. (2014). *Strategies in learning and using a second language* (2nd ed.). New York, NY: Taylor & Francis.

Dattalo, P. (2013). *Analysis of multiple dependent variables*. Oxford: OUP.

Di Stefano, C., Zhu, M., & Mîndrilă, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*. *14*(20). Retrieved from http://pareonline.net/getvn.asp?v=14&n=20

Dörnyei, Z. (2005). *Motivational strategies in the language classroom*. Cambridge: Cambridge University Press.

Dörnyei, Z. & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Oxford: Blackwell.

Hagell, P., Rosblom, T., & Palhagen, S. (2012). A Swedish version of the 16-item Parkinson fatigue scale (PFS-16). *Acta Neurol Scand*, *125*(4), 288–292. doi: 10.1111/j.1600-0404.2011.01560.x

Hirata, A. (2010). *An exploratory study of motivation and self–regulation learning in second language acquisition: Kanji learning as a task focused approach* (Unpublished MA thesis). Massey University, Manawatu.

Karami, H. (2012). The development and validation of a bilingual version of the vocabulary Size test. *RELC*. *43*(1), 53–67. doi: 10.1177/003368821 2439359

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Matsunaga, M. (2008). Item parceling in structural equation modeling: A primer. C*ommunication Methods and Measures*, 2(4), 260–293. doi: 10.1080/19312450802458935

Meada, A. W. & Kroustalis, C. M. (2006). Problems with item parceling for confirmatory factor analytic tests of measurement invariance. *Organizational Research Methods*, *9*(3), 369–403. doi: 10.1177/1094428105283384

Mizumoto, A. (2013). Enhancing self-efficacy in vocabulary learning: A self-regulated learning approach. *Vocabulary Learning and Instruction*, *2*(1), 15–24. doi: 10.7820/vli.v02.1.mizumoto

Mizumoto, A. & Takeuchi, O. (2012). Adaptation and validation of self-regulating capacity in vocabulary learning scale. *Applied Linguistics*, *33*(1), 83–91. doi: 10.1093/applin/amr044

Moir, J. & Nation, I. S. P. (2002). Learners' use of strategies for effective vocabulary learning. *Prospect*, *17*, 15–35.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

O'Neil, H. F., Jr. & Herl, H. E. (1998). *Reliability and validity of a trait measure of self-regulation*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Oxford, R. L. (2003). *Language learning styles and strategies: An overview*. Retrieved from hyxy.nankai.edu.cn/.../learning strategies by Oxford.pdf

Oxford, R. L. (2017). *Teaching and researching language learning strategies: Self-regulation in context* (3rd ed.). New York, NY: Routledge.

Pallant, J. (2013). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (5th ed.). Maidenhead: McGraw Hill.

Pintrich, P.R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*, 385–407. doi: 10.1007/s10648-004-0006-x

Schinka, J. A. & Velicer, W. F. (2003). *Handbook of psychology: Research methods in psychology* (Vol. 2). Hoboken, NJ: Wiley.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Trommsdorff, G. (2009), Culture and development of self-regulation. *Social and Personality Psychology Compass*, *3*, 687–701. doi:10.1111/j.1751-9004.2009.00209.x

Tseng, W., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, *27*(1), 78–102.

Tseng, W., Liu, H., & Nix, J. (2017). Self-regulation in language learning: Scale validation and gender effects. *Perceptual and Motor Skills*, *124*(2), 531–548. doi: 10.1177/0031512516684293

Tseng, W., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning: A structural equation modeling approach. *Language Learning*, *58*(2), 357–400. doi: 10.1111/j.1467-9922.2008.00444.x

Winne, P. H. & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 279–306). Hillsdale, NJ: Erlbaum.

Yeşilbursa, A. & Bilican, R. (2013). Validation of self-regulatory capacity for vocabulary learning scale in Turkish. *Procedia: Social-Behavioral Sciences*, *70*, 882–886. doi: 10.1016/j.sbspro.2013.01.134

Zeidner, M., Boekaerts, M., & Pintrich, P. R. (2000). Self-regulation: Directions and challenges for future research. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 750–768). San Diego, CA: Academic Press.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183. doi: 10.3102/0002831207312909

Zimmerman, B. J. & Camppillo, M. (2003). Motivating self-regulated problems solvers. In J. E. Davidson & R. J. Sternberg (Eds.), *The nature of problem solving* (pp. 233–262). New York, NY: Cambridge University Press.