Journal of Teaching Language Skills
(JTLS)
39(3.2), Fall 2020, pp. 111-142
DOI: 10.22099/jtls.2021.38867.2903

Online ISSN: 2717-1604
Print ISSN: 2008-8191

Research Paper

Development and Validation of a Measure of Self-Regulated Capacity in Learning the Grammar of English as a Foreign Language

Kioumars Razavipour [*] **Sara Safari Ardakani** [**]
Zohreh Gooniband Shooshtari [***]

Department of English Language and Literature, College of Letters and Humanities, Shahid Chamran University of Ahvaz, Iran

Abstract

Grammatical competence constitutes an important component of communicative ability, the acquisition of which takes sustained effort, resilience, and planning, otherwise known as the capacity for self-regulated learning. It follows that assessing the self-regulatory capacity in grammar learning (SRCgram) is of prime importance. This paper reports on the development and validation of a scale for measuring SRCgram. Focus group interviews were conducted with 26 participants and a pool of 52 items was created and piloted. Exploratory and confirmatory factor analyses were then conducted to examine the psychometric properties of the instrument. Preliminary fit indices, internal structure fit of the model, and overall model fit provided evidence for the validity of the scale. In addition, the SRCgram scale appeared to be unidimensional and of satisfactory reliability. Thus, SRCgram scale can be proposed as a diagnostic and self-assessment tool to be used by EFL teachers and

Received: 26/11/2020 Accepted: 18/01/2021

[*] Assistant Professor, Email: razavipur57@gmail.com, Corresponding Author
[**] M.A. Student, Email: sara.safari16@ymail.com
[***] Associate Professor, Email: shooshtari@yahoo.com

learners to diagnose, assess, and foster self-regulation in grammar learning.

***Keywords**:* Grammar Learning, Self-regulated Capacity, Validation, Factor Analysis, English Language

Unlike acquiring one's native language (L1), which seems to occur naturally with seemingly no intentional effort, learning a second or foreign language (L2), especially during adulthood, is known to be extremely challenging, especially when the target language is not used in the immediate social environment of the learner, which is the case in many countries in Kachru's expanding circle. Because of the slow pace of learning and frequent sliding back to one's previous level of mastery (Brown, 2000), it is all too easy for learners to lose motivation or experience feelings of failure and disappointment. Boredom and disappointment are even more likely to set in when it comes to learning the grammar of the target language, which is more abstract and less intuitively graspable than, say, vocabulary.

Grammatical competence is an important component of communicative competence (Purpura, 2013) as grammar is the common denominator of all the four language skills of reading, writing, speaking, and listening. In learning a second or foreign language, both teachers and learners highly value grammar (Sato & Oyanedel, 2019; Jean & Simard, 2011). Thornbury (1998, p. 19) underlines the importance of grammar by maintaining that "for as long as I have been teaching, grammar has never been anywhere *but* center stage" (italics in the original). Therefore, in all language teaching methods, there is a place for grammar teaching, though with varying degrees of emphasis (Andrews, 2008; Spinner, 2020).

Despite learners' crucial need to master grammar, acquiring L2 grammatical competence is known to be particularly challenging. Comeaux

and McDonald (2017) maintain that "one of the many challenges L2 earners face is acquiring the grammar and morphology of their new language" (p. 6). The challenges entailed in mastering the grammar of a foreign language demand for an efficient repertoire of strategies to help learners maintain their resilience and stay motivated in the face of setbacks they experience along the way. This is especially the case with adult learners who are no longer guided by the roadmap of a structured education within a formal educational setting. Roche (2018, p. 535) states that "one of the great attractions of adult learning is that we embark on the journey without quite knowing the destination". For language learners to make the most of the noted attractions and not to be negatively affected by this lack of a clear destination, they need to be equipped with the right strategic competence (Bachman, 1990).

Strategic language learning is conventionally assessed using self-report questionnaires, which are based on the premise that strategies are indicative of an underlying trait. This is so because questionnaire "items ask respondents to generalize their actions across situations rather than referencing singular and specific learning events" (Tseng, Dörnyei, & Schmitt, 2006, p. 82). In spite of this assumption, most self-report instruments for measuring strategic learning of an L2 consist of items that target discrete language learning strategies and specific strategic behaviors (Tseng et al, 2006), rendering it difficult to relate the items in these measures to a coherent underlying theoretical model (Dörnyei & Skehan, 2003). In addition, when it comes to interpreting the results of self-report questionnaires, the going-togetherness of individual item scores with that of the entire scale cannot be assumed (Tseng et al., 2006). Designing the questionnaire items on the basis of the frequency of the behaviors is yet another drawback of the conventional self-report questionnaires. learning strategy theory indicates that strategy use is a matter of quality of the strategies used rather than their quantity; therefore, more

frequent use of strategies does not necessarily translate into more effective language learning, nor is low reported use of strategies a clear sign of ineffective learning.

Taking these issues into account in the light of the noted theoretical and measurement pitfalls in language learning strategy measures, there is a need to conceptualize, develop, and test an instrument targeting the learner's trait of self-regulatory capacity in learning English grammar in which the self-report items constitute the general declaration and conditional relations. Since the notion of learning strategies is still theoretically problematic, Dörnyei (2008) self-regulatory system, consisting of five facets, can be an appropriate alternative theoretical frame of reference based on which a self-regulatory inventory aligned with grammar can be developed. This theory has already informed the successful development and validation of a scale for the measurement of self-regulated learning of vocabulary (Tseng et al., 2006). Using the same theoretical framework and building on the work of Tseng and his colleagues, the present study aimed to devise, pilot, and validate a measure of SRCgram. Given the benefits that mastery of grammar accrues to the language learner (Derewianka, 2019), the developed scale can be used for diagnosis, placement, and selection functions for both pedagogical and assessment purposes. In addition, adult learners who continue their language learning independently will benefit from using the noted scale in diagnosing what is going wrong in their grammar learning, which would, in turn, help them stay on track and maintain their commitment to learning an additional language.

Review of Literature

In the language learning literature, to refer to learners' own conscious decisions as to how to plan, execute prioritize, and monitor their language

learning, various overlapping terminology is often used. Most frequently used among such terms are independent learning, self-regulated learning, self-directed learning, self-management, autonomous learning, and language learning strategies (see Cohen, 2014). However, the latter term, learning strategies, is more commonly used in the literature. Simply put, language learning strategies (LLS) are "actions chosen by learners for the purpose of learning a language" (Griffiths 2019, p. 2).

Whereas acquiring an L1 occurs naturally and apparently uniformly for almost all children, learning an L2 is a far cry from a smooth, effortless, and uniform process. In the language learning literature, blame was, for a while, placed on language teaching methods for the failure of adult L2 learners (Macaro, 2001), on the grounds that teaching methods and textbooks attended exclusively to grammar and translation. In reaction to the heavy emphasis on translation and grammar, the Direct Method of language teaching gained popularity, to be replaced later with the Audio-lingual method, which in turn gave way to communicative and task-based language teaching methods. Nevertheless, it has been observed that regardless of variations in the content or methods of language teaching, some learners outperform others in learning a second language (Griffiths 2015; Macaro 2001; Rubin 1975; Stern, 1975). This observation led scholars to conclude that there must be some differences between the good language learners' approaches to language learning and those of the less successful learners. It was believed that if the learning habits and strategies that set the good language learner apart from the less successful ones could be identified, they could be imparted to the less successful learners, hence rendering successful learning of a second language a possibility for all language learners (Oxford 2002; Robin 1975). In simple terms, the idea of learning strategies was in fact about teaching learners how to fish for themselves rather than giving fish to them. It was about learning to learn

(Macaro 2001). If learners learn how to learn a language, they would be autonomous and can sustain their momentum despite unfavorable teaching and learning circumstances. This would in turn guarantee access to a second language for a lifetime (Cohen, 2014).

This recognition of the language learning strategies and their potential to build learner autonomy (Hsiao & Oxford, 2002) generated a number of theoretical and empirical taxonomies of language learner strategies. Rubin's (1981) divided strategies into direct and indirect strategies. Oxford (1990) modified the direct/indirect classification. Memory, compensation, and cognitive strategies were subsumed under direct strategies, and the indirect strategies were subdivided into social, affective, and metacognitive ones. In O'malley and Chamot (1990) model, language learner strategies were grouped into socio-affective, cognitive, and metacognitive. The next phase in research on language learning strategies focused on strategies for specific areas or skills of language such as vocabulary learning (Taka 2008) and test-taking (Cohen 2014).

Despite the surge of interest in language learner strategies, several issues have been, and continue to be, open to debate. Quite early on, definitional fuzziness and construct elusiveness of language learning strategies concerned scholars (Griffiths, 2019). As such, it is still not clear whether a strategy refers to observable behavior, an invisible mental process, or both (Tseng et al, 2006). Research in LLS has been criticized on the methodological front as well (Reid 1990). Reid warns against the use of surveys on learning styles and strategies whose validity and consistency for the target context has not been examined.

More importantly, the theoretical genealogy of research in LLS continues to be controversial. Some scholars maintain that this strand of research is atheoretical (Dörnyei and Skehan, 2003; Tseng et al, 2006). That is, it is not

clear to which theory of language learning LLS research is aligned with. Recently, proponents have tried to situate LLS research within some theoretical perspective. For example, Griffiths (2019) maintains that cognitivism, humanism, and sociocultural theories can all be mobilized to afford SSL research with some theoretical anchorage. Yet, to what extent such grand, diverse theories can provide a coherent theoretical framework remains open to debate, to say the least.

As to research on autonomous, strategic learning of grammar, a similar reliance on inventories of discrete learning strategies has been common. Usually, scholars use the same taxonomies of general LLS to assess strategic learning of grammar (Alsied, Ibrahim, and Pathan 2018; Bayou 2015; Gürata 2008; Pawlak, 2012; Zekrati 2017). Hence, the same issues pointed out with regard to assessing general SSL linger in the assessment of autonomy in grammar learning. To reiterate, the measurement of strategic language learning suffers from both theoretical conceptualization and methodological issues. On the theoretical front, lack of coherent theory limits the interpretability of numerical data gleaned from SSL inventories. For the purpose of measurement, the theory is of critical importance because as Messick (1989) maintains, the validity of a measuring instrument is the extent to which empirical evidence and theoretical rational point to the defensibility of inferences and actions that are to be made based on scores from a measuring instrument. Therefore, measuring instruments (i.e., tests, surveys, questionnaires, etc.) with questionable theoretical foundations are not likely to result in valid inferences or actions.

On the methodological front, the LLS measures do not satisfy the condition of additivity (see Heene, 2013), implying that a higher frequency of strategy use does not necessarily translate into more successful, strategic learning. In other words, the causal relationship between scores from the noted

inventories and higher levels of the autonomy construct is not satisfied (Borsboom 2005; Heene 2013).

For the noted problems, some scholars have suggested abandoning LLS research and shifting to self-regulated learning, which has a more robust theoretical foundation and holds more promise in assisting language learners 'lifetime learning. Accordingly, Tseng et al. (2006) pioneered the use of self-regulation theory for the measurement of self-regulated, autonomous learning of vocabulary. Our study builds on the work of Tseng et al. (2006) in designing and validating a scale of measuring SRCgram for adult learners. The advantage of drawing on self-regulation theory rather than on conventional strategy research is that "rather than focusing on the outcomes of strategic learning (i.e., the actual strategies and techniques the learners apply to enhance their own learning)", the self-regulation conceptual approach, "highlights the importance of the learners" innate self-regulatory capacity that fuels their efforts to search for and then apply personalized strategic learning mechanisms." (Tseng et al., 2006, p. 79). The self-regulatory model underpinning the design and development of SRCgram comprises five facets, namely, commitment control, metacognitive control, satiation control, emotional control, and environment control (Dörnyei, 2008), each of which is briefly defined below. Commitment control helps learners maintain their motivation to pursue their original goal of grammar learning. Metacognitive control assists learners in overseeing their concentration and resisting procrastination. Satiation control helps learners avoid the boredom that might occasionally set in the course of their grammar learning. Emotional control is the facet that helps learners counter disruptive emotions by mustering feelings that would help them stay on track. Finally, environmental control "helps to eliminate negative environmental influences and to exploit positive environmental influences by making the environment an ally in the pursuit of

a difficult goal" (Tseng et al. 2006, p. 86). In sum, the above review reveals that though several attempts have been made to apply the theory of self-regulation from educational psychology to language learning, to our knowledge, it has not been studied in relation to learning grammar. More specifically, no measure has been developed to assess self-regulated capacity in grammar learning, a gap this study intends to narrow.

The remainder of this paper reports on the methods of the study, followed by analytic procedures and findings. The paper concludes with conclusions, implications of the study, and a few ideas for further inquiry.

The Study

This study was conducted in the EFL context of Iran, where learners have very few if any, opportunities for authentic interaction with native speakers of English or for putting their English knowledge to everyday use. As such, they have to be highly motivated and self-regulated to maintain their passion and interest, especially in learning grammar which is probably the least interesting aspect of language learning. A total of 368 undergraduate students majoring in English in two state universities in the Southwest of Iran contributed data to this study. To generate the questionnaire items, twenty-six participants took part in focus group interviews conducted to generate items. These items were then piloted with 114 students, 35 male and 79 female. Another 228 students, 72 male and 156 female, participated in the main study phase. Consistent with the student populations in the English language departments of the country, which are often female-dominated, females constituted two-thirds of the participants.

Item Generation and Design of the Scale

To generate the item pool for the measurement of learning scales, scholars recommend involving learners in the process (Dörnyei & Taguchi, 2009). Along with such advice, three focus group interviews were held with 26 participants by the second author. The interview prompts were informed by the theory of self-regulated learning (elaborated on above) comprising of five control mechanisms. In other words, these mechanisms served as a rough blueprint for the generation of interview questions (see Phakiti, 2021). To allow for deeper and more elaborate insights into their self-regulatory process in grammar learning, the interviews were conducted in Persian. The interview questions were derived based on the theoretical model of self-regulated language learning proposed by Dörnyei (2008). Insights from interviews coupled with our review of the literature led to the generation of 52 Likert type items, on the five subscales of self-regulated language learning discussed earlier.

To neutralize the likely confounding influence of differential English proficiency on the part of the participants, the items were written in Persian, the official language of Iran. The pilot version of the scale was then administered to the pilot study participants with the second author being present at the research site to explain the study purpose, to assure respondents of the confidentiality of the data, to offer clarifications if needed, and to benefit from the participants' possible occasional feedback on items.

To identify items with poor performance, Extreme Group Method and Item-Total Correlation were carried out (See Kaplan & Saccuzzo, 2004). Five items showed poor item-total correlations, which were dropped from the instrument.

For Extreme Group Method Analysis, items that could discriminate between the upper 33% and the lower 33% of the participants were retained

(Brown, 2005). Given the ordinal nature of individual Likert items, the Mann-Whitney U test was used for doing extreme group analysis (Pallant, 2013). Three items that failed to discriminate significantly between the upper and the lower total scores on the scale were eliminated.

The next procedure used to assess the effectiveness of the items was Corrected item-total Correlation, which examines the degree to which each item correlates with the total subscale score with which it is aligned. Corrected Item-Total Correlations smaller than .3 are considered poor (Pallant, 2013) and are candidates for deletion. Accordingly, eight items that failed to reach the cut-off point of .3 were removed from the remaining 44 items. Hence, we were left with 36 items at this stage. The instrument was then piloted and further item and scale evaluation were carried out.

In specific, exploratory factor analysis was first conducted. Prior to that, the general parametric requirements were checked and the factorability of the data was examined using KMO and Bartlett's test of sphericity. Given our realist assumptions concerning the psychometric reality of self-regulation and its control dimensions, EFA was carried out using Principal Axis Factoring with direct oblimin rotation. Eleven items with cross-loadings on multiple factors or with loadings less than a threshold of .4 (Pallant, 2013) emerged and subsequently were eliminated. Thus, we ended up with 25 items (see Appendix 1), five for each subscale of the self-regulatory model explained earlier. A preliminary reliability check with pilot study data suggested that individual subscales and the scale as a whole enjoyed acceptable indices (see Table 1). It should be noted that factors were named based on the content of the items that loaded on them.

Table 1.

Internal Consistency Reliability Indexes

Subscale	Alpha	Number of items
Commitment Control	.795	5
Metacognitive Control	.705	5
Satiation Control	.735	5
Emotional Control	.797	5
Environmental Control	.694	5
Scale	.925	25

As can be seen, for all subscales, except for environmental control, which is borderline, other subscales meet the consistency threshold required. The reliability of the scale as a whole is also .925, which is close to ideal. In the measurement literature, it is held that for any measure to be valid, it has to be reliable (Bachman & Palmer, 2010; Sarstedt & Mooi, 2014). Therefore, once we were assured of the reliability of the SRCgram and had tentative evidence concerning the distinct nature of the five factors underlying it, we moved to the next phase and administered the scale to the main study participants. The new data were then analyzed using Confirmatory Factor Analysis, which entails the following stages: specification, identification, estimation, evaluation, and modification. The model of self-regulated learning was specified to include five control components as indicated in the literature and emerged in EFA, as noted above. For model identification, we used the t-rule (see Byrne, 2013) to ensure that the model was over-identified and thus estimable. Details regarding model estimation and evaluation are given in the next section.

Results

Before conducting further statistical analyses, the reliability of subscales and the scale as a whole was checked again using the main study data. Table

2 gives the Cronbach's alpha indices of reliability for the scale in the main study. This provides evidence to the measurement invariance (Markus & Borsboom, 2013) of the scale both across time and participants. Simply put, measurement invariance requires that the properties of a measuring tool not fluctuate with the uses to which it is put.

Table 2.
Reliability Analysis Based on Main Study Data

Subscale	Alpha	Number of items
Commitment Control	.762	5
Metacognitive Control	.690	5
Satiation Control	.774	5
Emotional Control	.747	5
Environmental Control	.715	5
Scale	.922	25

Table 2 presents the indices of internal consistency for the subscales and the scale as a whole. As we see, the index for all subscales exceeds the threshold of .7 (DeVellis, 2016) except for the Metacognitive Control subscale that is a borderline case. The internal consistency coefficient of the whole scale was .922, which together with the obtained values for subscales can be taken as evidence of the satisfactory reliability of the scale. Accordingly, the SRCgram appeared to satisfy both temporal and sample aspects of measurement invariance, for it demonstrated good reliability across two occasions with two different samples of participants.

Confirmatory Factor Analysis (CFA)

As a powerful statistical tool to gauge the extent to which the underlying theory is borne out in statistical data, CFA provides evidence of construct validity (Byrne, 2013; Phakiti, 2018). Therefore, we conducted CFA to further

examine the construct validity of SRCgram. To do so, following Tseng et al. (2006) and Bagozzi and Yi (1988), to assess the postulated model against empirical data, preliminary fit indexes and global fit criteria were used. The internal structure of the model in relation to its constitutive components was also assessed.

As for preliminary fit criteria, factor loadings, correlations among variables (to assess divergent validity), and standard errors must be checked (Tseng et al., 2006). Ideas regarding the cut-off point loading for items in CFA diverge. Kline (2005) and Qasemi (2013) suggest that values between 0.3 to 0.6 are acceptable and those larger than 0.6 are considered high.

Table 3.
CFA Factor Loadings of Items

Self-Regulation Dimensions	Item No.	Standardized Regression Weights	Status
Commitment control	51	.738	Good
	23	.593	Good
	31	.615	Good
	38	.584	Good
	30	.575	Good
Emotional Control	52	.731	Good
	15	.600	Good
	50	.613	Good
	37	.676	Good
	46	.481	Good
Environmental Control	33	.677	Good
	22	.527	Good
	4	.552	Good
	10	.452	Good
	13	.668	Good
	45	.452	Good
	2	.583	Good

Self-Regulation Dimensions	Item No.	Standardized Regression Weights	Status
Metacognitive Control	36	.606	Good
	21	.530	Good
	48	.439	Good
Satiation Control	32	.719	Good
	29	.687	Good
	40	.661	Good
	7	.555	Good
	43	.566	Good

As both Table 3 and Figure 1 illustrate, all factor loadings fall within the acceptable range noted above.

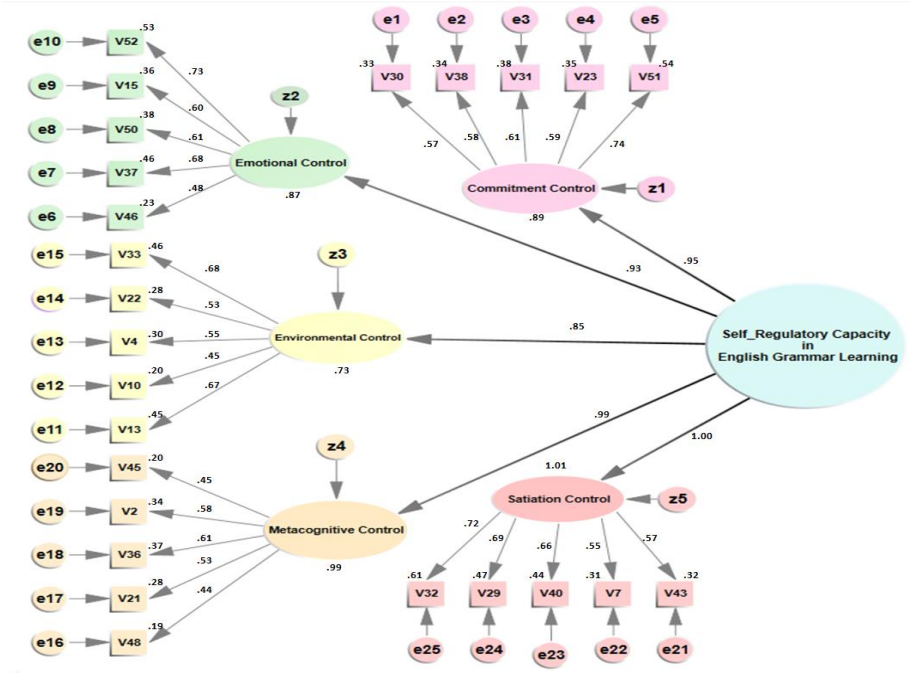

Figure 1.
SRCgram Confirmatory Factor Analysis Model (Standardized Values)

The next preliminary fit criterion used was correlations among subscales. Moderate correlations among subscales of a measure are considered evidence for the convergent validity of the scale; on the other hand, too high correlations among subscales are indicative of a lack of divergent validity (Pallant, 2013).

Table 4.
Correlations Among Subscales

	Satiation Control	Metacognitive Control	Environmental Control	Commitment control	Emotional Control
Satiation Control	1.000				
Metacognitive Control	.693	1.000			
Environmental Control	.639	.512	1.000		
Commitment control	.713	.629	.577	1.000	
Emotional Control	.738	.637	.533	.707	1.000

Table 4 shows moderate correlations among the subscales, lending support to the distinctiveness and divergent validity of the subscales.

The scale was further evaluated using overall model fit criteria, summarized in Table 5.

Table 5.
The goodness of Fit Criteria

Fit indices	**Reported fit statistic**	
Chi-square	588.070	
Relative Chi-square	2.146	<5
RMSEA	.071	<.08
CFI	.948	>.9
NFI	.751	>.9

Fit indices	Reported fit statistic	
GFI	.825	>.9
AGFI	.729	>.9
IFI	.850	>.9
TLI	.833	>.9
CN	122> 68 at .05; 129> 73 at .01	

The insignificance value of the chi-square statistic is a requirement for a model to be acceptable. In other words, for a model to demonstrate a good fit, an insignificant result at 0.05 must be obtained (Barrett, 2007), and that is why some consider the Chi-Square statistic as a 'badness of fit' index (Liu, 2015). The value of 588.070, under CMIN, was significant for the *p*-value of .000 and the model was rejected. Not much can be read into this because the Chi-square statistic is a test of significance which, like other significance tests, varies depending on the sample size (Byrne, 2013). Accordingly, with large sample sizes, a model is almost always rejected by the Chi-square statistic. Sample sizes larger than 200 are considered to compromise the dependability of the chi-square test as a goodness of fit criterion (Bagozzi & Yi, 1988). Considering the fact that we had 229 participants in this study, CMIN could not serve as a reliable fit index.

Various values have been proposed regarding the criterion for acceptance of relative chi-square, with some advising a value less than 2 (Byrne, 2013) and others recommending a value less than five (Lomax & Schumacker, 2004). This ratio was 2.146 for our model, implying a perfect match as recommended in the literature (Jöreskog & Sörbom, 1993; Kline, 2005).

The RMSEA index in Table 5 is 0.071. Some scholars consider acceptable an RMSEA value smaller than .08 (Phakiti, 2018); others believe that it should ideally be smaller than .05 (Byrne, 2013). Accordingly, the

RMSEA index we obtained suggests an acceptable fit between the hypothesized model and the observed data.

Two commonly used comparative fit indices used are CFI and NFI. The CFI value reported in Table 5 is 0.948. CFI values of 0.90 or higher suggest very good fit of the model to the data (Byrne, 2013). Therefore, we can safely claim that the CFI index of the model is indicative of a good model fit. Likewise, for IFI, values closer to one are regarded as the more acceptable (Qasemi, 2013), though other scholars suggest higher values. In this study, the IFI index was 0.850. Based on the criterion suggested by Qasemi (2013), this index is a tolerable index of fit. That said, according to Bentler (1990, as cited in Byrne (2013), of the two comparative fit indices, "CFI should be the index of choice" (p. 79).

Both GFI and AGFI indices range from zero to one (Hooper, Coughlan, & Mullen, 2008), with values close to 1.00 being indicative of a good fit (Byrne, 2013). The values of 0.825 and 0.792 were respectively found for GFI and AGFI statistics, suggesting that they are only tolerable, possibly by the least stringent criteria. Due to its sensitivity, GFI index "has become less popular in recent years and it has even been recommended that this index should not be used" (Hooper, Coughlan, & Mullen, 2008, p. 54). In addition, as AGFI tends to increase with sample size (Hooper et al., 2008), this fit index is not often relied upon as a standalone index of fit.

CN values in excess of 68 at 0.05 and 73 at 0.01 signify a proper fit between a model and empirical data (Tseng et al., 2006). The reported CN statistic in the present study, 122 at 0.05 and 129 at 0.01, respectively, which are greater than the said thresholds, could be interpreted that the employed sample size (229) in the model was satisfactory. In sum, it can be concluded that, except for TLI, the goodness of fit criteria reported in Table 5 lend support to the adequacy of the model. Taken together, various fit indices

suggest that though the fit between data and the model is not perfect, it can be considered as acceptable.

As the final step in the evaluation of the CFA model, the fit of the internal structure of the model was examined. In so doing, three reliability indexes, as recommended in Bagozzi and Yi (1988), were examined. These reliability types were the following: reliability of individual items, reliability of the scale as a whole, and the average extracted variance for each subscale.

Table 6.
Individual Item Reliability of Each subscale

Self-Regulation Dimensions	Individual Item Reliability	Evaluation
Commitment Control	0.618	Good
Emotional Control	0.62	Good
Environmental Control	0.576	Good
Satiation Control	0.638	Good
Metacognitive Control	0.522	Good

The recommended threshold value for Individual item reliability is 0.50; that is, values above .5 meet the reliability criterion (Tseng et al., 2006). Table 6 summarizes the results of individual item reliability of each of the self-regulatory capacity subscales. All the five subscales have values larger than the recommended cut-point (0.50) and thus are well within the satisfactory range.

The last index of the internal structure fit we examined was composite reliability statistic, which refers to the overall reliability of the scale. A value of 0.60 is suggested as the cut-off point (Tseng et al., 2006). This value was computed via Composite Reliability Calculator, an application provided in the Statistical Mind website (http://www.thestatisticalmind.com/composite-reliability). The resultant value of 0.933 exceeded the recommended

threshold, attesting to the satisfactory reliability of the whole scale. The average extracted variance calculated value in this study was 0.5948, which was greater than 0.50, pointing to the good internal structure of the measure.

In sum, preliminary fit indexes, global fit criteria and the internal structure of the model lend support to the psychometric soundness of SRCgram.

Discussion

Mastering the grammar of a second language is a long journey demanding a firm determination backed by proper self-regulated mechanisms. Without proper tools to assess the extent to which learners at every stage of their development can harness the self-regulatory resources for achieving their grammar learning goals, it is difficult for teachers to discern where on the continuum of intention to goals learners need support and guidance. In this study, we aimed to develop and validate a practical scale for the assessment of self-regulated capacity in learning the grammar of English as a foreign language. To overcome the limitations of strategy-based inventories, which often describe specific behaviors rather than generic tendencies holding across contexts, we built on a model of self-regulated learning to design a psychometrically sound instrument for assessing SRCgram. In a three-phase study, an initial item pool was created based on the past literature and focus group interviews, and the psychometrically sound items were identified through a series of statistical analyses such as item analysis, reliability analysis, and exploratory factor analysis. CFA was then conducted to assess the extent of fit between empirical data and the postulated self-regulated model of English grammar learning. Using a series of fit criteria, the designed SRCgram scale appeared to be of acceptable reliability and validity.

Though SRCgram was intended for assessing self-regulated grammar learning, as the name suggests, its underlying five-factor conceptualization lends support to the self-regulation model that informed the study conducted by Tseng et al. (2006). Consistent with Tseng et al. (2006), we found that the five control dimensions of *commitment*, *emotion*, *environment*, *satiation*, and *metacognition* account for a good deal of variation in the data. On the other hand, our findings are discrepant with Mizumoto and Takeuchi (2013), who replicated Tseng et al's study in Japan. Contrary to their expectation, fit indices from CFA were rather poor for a five-factor solution. Alternatively, they dropped the satiation and commitment factors from the original SRCvoc model and added procrastination as a latent factor that must negatively correlate with self-regulation. This structural model turned out to show good model fit indices as well as providing evidence of divergent validity. Though the researchers did exercise caution in translating the scale, it is likely that some of the original properties of the measure were lost in the process of translating it into Japanese (see Drasgow & Probst, 2005). It might also be due to the influence that the social environment exerts on self-regulated learning because "self-regulatory skills are acquired through social modeling, social guidance and feedback, and social collaboration" (McInerney & King, 2011, p. 485).

Self-regulation theory has informed the development of similar measures in relation to other areas in L2 learning. Hu and Gao developed and validated a questionnaire to assess self-regulated learning of writing in EFL. Using confirmatory factor analysis, they discovered that the self-regulatory model that best fitted the data had a hierarchical structure comprising of nine factors and self-regulation as the higher-order factor accounting for the correlations among the first-order factors. Perhaps the divergent findings have to do with the nature of writing which is a more complex social activity

(White, 2019) compared with grammar. As such, self-regulated learning of writing might entail more nuanced models of self-regulation and metacognition than that of grammar.

It seems that the available literature has addressed only the explanation inference in an overall evaluative argument for measures of self-regulated language learning. Hence, robust validation programs may be worth pursuing to scrutinize the evaluation, generalizability, extrapolation, and utilization inferences. More fundamentally, we must not lose sight of the fact that the entire discourse of latent trait measurement, including factor analysis, is no more than a discourse common among an academic tribe (Hyland, 2006) and we must beware of the trap of committing the fallacy of nominal realism (McGrane & Maul, 2020) by thinking that the mere existence of a name necessitates that it represents something of ontological reality. Hence, the discourse of modeling latent traits should not beguile us into conflating the map with the world (Maraun, 2010). Therefore, no matter how perfect the fit indices are, the profound questions of exactly what self-regulated learning is and whether it exists at all must not be taken for granted or considered secondary.

A final issue that must be borne in mind in interpreting the findings of this study is the nature of the construct of grammar. Although the participants of this study seemed to have a tacitly clear definition of what grammar is, as none questioned the scope of the construct, there is considerable controversy surrounding the construct of grammar (Spinner, 2021). In particular, the scope of grammar, its boundaries with other components of communicative ability such as pragmatics and lexicon remain controversial (Purpura, 2014; Spinner, 20121). Therefore, it should be emphasized that given the background of the participants of this study, who were English major students required to take credit courses in grammar and the nature of textbooks that are often used in

such courses, perhaps it is safe to claim that the scale developed in this study taps into self-regulated learning of grammar in its traditional sense, which is about language forms and structures.

Conclusion

One pedagogical lesson to take away from the findings of this study is that in this era of rapid change, spoon-feeding students in writing and grammar courses can no longer be a viable solution to the challenges of learning English. However, language teachers and educational policymakers should aspire to prepare learners to embark on their own individualized learning paths, even long after they cease attending language courses. To do so, learners must be trained to exercise control over their learning habits, motivational states, saturation feelings, and environmental disturbances. SRCgram can serve to diagnose what aspects of learners' self-regulated learning need improvement or change for more sustained learning.

Several limitations must be kept in sight regarding the findings of the current study. The first limitation has to do with the temporal characteristic of this research project; it was cross-sectional in nature. Given that self-regulatory capacity, like other psychological constructs, undergoes change over time, longitudinal studies hold the promise to further our understanding of self-regulation in grammar learning under various circumstances.

Further, as noted above, past research suggests that absolute universality is difficult to maintain when it comes to the validity of measurements in the humanities and social science (Kane, 2013; Weir, 2005). The present study was conducted in the context of Iranian EFL learners. Because of the dynamic nature of self-regulation, its structure may vary across contexts (Mizumoto & Takeuchi, 2012). In addition, the size of the sample precluded using other

analytical procedures like Rasch Analysis. Hence, the validity of SRCgram can be further examined and enhanced using item response theory.

The other limitation was the adoption of a self-report questionnaire as the main data elicitation tool, which, despite its ubiquity and popularity, may not capture all the nuances of self-regulated learning because of the high degree of awareness and introspection that is demanded of the respondents. Although self-report instruments are widely used in the study of SRL, triangulation through other methods of data collection such as stimulated recall can further enhance the credibility of the findings. Another line of inquiry is to examine the predictive validity of the SRCgram to see the extent to which variations in scores on SRCgram are associated with similar patterns in actual grammar learning. Finally, the choices that language learners make in the course of their learning are influenced, at least partly, by the surrounding sociocultural milieu (Griffiths, 2019; Lantolf, 2000; Reid, 1990). Therefore, SRGgram must be examined for its validity across diverse cultural and educational settings.

Despite the noted limitations and our acknowledgment that it awaits further validation, we believe that SCRgram is a useful tool in helping learners self-regulate their learning of English grammar at a time where English has become the lingua franca of the world of business, technology, and academic publication (Jenkins, 2013; Seidlhofer, 2005).

References

Alsied, S. M., Ibrahim, N. W., & Pathan, M. M. (2018). The use of grammar learning strategies by Libyan EFL Learners at Sebha University. *ASIAN TEFL Journal of Language Teaching and Applied Linguistics*,*3*(1), 37-51.

Andrews, S. (2008). *Teacher language awareness*. Oxford: Oxford University Press.

Bachman, L. F. (1990). *Fundamental considerations in language testing*: Oxford university press.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*: Oxford University Press.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the academy of marketing science*, *16*(1), 74-94.

Barrett, P. (2007). Structural equation modeling: Adjudging model fit. *Personality and Individual differences*, *42*(5), 815-824.

Bayou, Y. (2015). *Grammar Learning strategies use of grade 11 students at Medhanealem preparatory school: Gender in focus.* Addis Ababa University.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238-246

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588.

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

Brown, H. D. (2000). *Principles of language learning and teaching*. New York: Longman.

Brown, J. D. (2005). *Testing in language programs: a comprehensive guide to English language assessment*. New York: McGraw-Hill College.

Sato, M., & Oyanedel, J. C. (2019). "I think that is a better way to teach but…": EFL teachers' conflicting beliefs about grammar teaching. *System, 84*, 110-122.

Byrne, B. M. (2013). *Structural equation modeling with EQS: Basic concepts, applications, and programming*. London: Rutledge.

Cohen, A. D. (2014). *Strategies in learning and using a second language*. New York: Routledge.

Comeaux, I., & McDonald, J. L. (2017). Determining the effectiveness of visual input enhancement across multiple linguistic cues. *Language Learning*, *68*(1), 5-45.

Drasgow, F., & Probst, T. M. (2005). The psychometrics of adaptation: Evaluating measurement equivalence across languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (eds). *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). New Jersey: Lawrence Erlbaum.

Derewianka, B. (2019). A relevant pedagogic grammar for today's classrooms. In X. Gao (ed). *Second handbook of English language teaching* (pp. 821- 845). Switzerland: Springer.

DeVellis, R. F. (2016). Scale development: Theory and applications. CA: Sage publications.

Dörnyei, Z. (2008). *Motivation strategies in the language classroom*. Stuttgart: Ernst Klett Sprachen.

D¨ornyei, 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). Malden, MA: Blackwell.

Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in second language research: Construction, administration, and processing*. London: Routledge.

Griffiths, C. (2015). What have we learned from 'good language learners'? *ELT Journal*, *69*(4), 425-433.

Griffiths, C. (2019). Language learning strategies: Is the baby still in the bathwater? *Applied Linguistics*. Electronic publication. Published online: 11 January 2019 (doi: 10.1093/applin/amy024).

Gürata, A. (2008). *The grammar learning strategies employed by Turkish university preparatory school EFL students (*M.A thesis). Bilkent University.

Heene, M. (2013). Additive conjoint measurement and the resistance toward falsifiability in psychology. *Frontiers in psychology*, *4*, 246.

Hsiao, T. Y., & Oxford, R. L. (2002). Comparing theories of language learning strategies: A confirmatory factor analysis. *The Modern language journal*, *86*(3), 368-383.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modeling: Guidelines for determining model fit. *Electronic journal of business research methods*, *6*(1), 53-60.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Hyland, K. (2006). *English for academic purposes*. Routledge.

Jean, G., & Simard, D. (2011). Grammar teaching and learning in L2: Necessary, but boring? *Foreign language annals*, *44*(3), 467-494.

Jenkins, J. (2013). *English as a lingua franca in the international university: The politics of academic English language policy*. London: Routledge.

Jöreskog, K. G., & Sörbom, D. (1993). LISREL 8: *Structural equation modeling with the SIMPLIS command language*. Scientific Software International.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1-73.

Kaplan, R. M., & Saccuzzo, D. P. (2004). *Psychological testing: Principles, applications, and issues*. Toronto: Nelson Education.

Kline, T. J. (2005). *Psychological testing: A practical approach to design and evaluation*. New York: Sage Publications.

Lantolf, J. P. (2000). *Sociocultural theory and second language learning*. Oxford: Oxford University Press.

Liu, A. (2015). *Structural Equation Modeling and Latent Variable Approaches.* John Wiley & Sons, Inc.

Lomax, R. G., & Schumacker, R. E. (2004). *A beginner's guide to structural equation modeling*. London: Psychology Press.

Maraun, M. D. (1996). Meaning and mythology in the factor analysis model. *Multivariate Behavioral Research*, *31*(4), 603-616.

McInerney, D. M. & King, R. B. (2011). In B. J. Zimmerman & D. Schunk, D. H. (Eds.). *Handbook of self-regulation of learning and performance (pp. 485-502)*. New York: Routledge.

Macaro, E. (2001). *Learning strategies in foreign and second language classrooms: The role of learner strategies*. London: Continuum International Publishing.

Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York: Routledge.

McGrane, J. A., & Maul, A. (2020). The human sciences, models and metrological mythology. *Measurement*, *152*, 107346.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement*. New York: Macmillan.

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

O'malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge university press.

Oxford, R. L. (2002). Language learning strategies in a nutshell: Update and ESL suggestions. *Methodology in language teaching: An anthology of current practice*, *4*(3), 124-132.

Pallant, J. (2013). *SPSS survival manual: A step by step guide to data analysis using SPSS*. Berkshire: McGraw-Hill Education

Pawlak, M. (2012). Instructional mode and the use of grammar learning strategies. In M. Pawlak (ed). *New perspectives on individual differences in language learning and teaching* (pp. 263-287). Berlin: Springer.

Phakit, A. (2021). Likert-type Scale Construction. In P. Winke, & T. Brunfaut (eds). *The Routledge handbook of second language acquisition and language testing* (pp. 102-114). New York: Routledge.

Phakiti, A. (2018). Confirmatory factor analysis and structural equation modeling. In A. Phakiti, P. De Costa, L. Plonsky, & S. Starfield. (eds). *The Palgrave handbook of applied linguistics research methodology* (pp. 459-500): Springer.

Purpura, J. E. (2013). Assessing grammar. In A. Kunnan (ed). *The companion to language assessment* (pp. 100-124). New York: John Wiley & Sons.

Qasemi, V. (2013). *Structural equation modeling in social research using Amos Graphics*. Tehran: Jame'eh Shenasan Publications.

Reid, J. (1990). The dirty laundry of ESL survey research. *TESOL Quarterly*, *24*(2), 323-338.

Roche, S. (2018). What's the score? Assessing the impacts and outcomes of lifelong learning. *International Review of Education*, *64*, 535-542.

Rubin, J. (1975). What the" good language learner" can teach us. *TESOL Quarterly*, *9*, 41-51.

Rubin, J. (1981). Study of cognitive processes in second language learning1. *Applied Linguistics*, *2*(2), 117-131.

Sarstedt, M., & Mooi, E. (2014). *A concise guide to market research. The Process, Data, and Methods Using IBM SPSS Statistics*. Berlin: Springer.

Seidlhofer, B. (2005). English as a lingua franca. *ELT Journal*, *59*(4), 339-341.

Stern, H. H. (1975). What can we learn from the good language learner? *Canadian Modern Language Review*, *31*(4), 304-319.

Spinner, P. (2021). Measuring grammar. In P. Winke, & T. Brunfaut (eds). *The Routledge handbook of second language acquisition and language testing* (pp. 233-242). New York: Routledge.

Taka, V. P. (2008). *Vocabulary learning strategies and foreign language acquisition*. Clevendon: Multilingual matters.

Teng, L. S, Zhang L. J. (2016). A questionnaire-based validation of multidimensional models of self-regulated learning strategies. *Modern Language Journal*, *100* (3):674–701

Thornbury, S. (1998). Comments on Marianne Celce-Murcia, Zoltán Dörnyei, and Sarah Thurrell's "Direct Approaches in L2 Instruction: A Turning Point in Communicative Language Teaching?": A Reader Reacts…. *TESOL Quarterly, 32*(1), 109-116.

Tseng, W.-T., 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.

Weir, C. J. (2005). *Language testing and validation*. Hampshire: Palgrave McMillan.

White, E. (2019). (Re) visiting twenty-five years of writing assessment. *Assessing Writing*, *42*, 100419.

Zekrati, S. (2017). The relationship between grammar learning strategy use and language achievement of Iranian high school EFL learners. *Indonesian EFL Journal*, *3*(2), 129-138.

Appendix. *The SRCgram Items*

Item No.		*Strongly agree*	*Agree*	*Slightly agree*	*Slightly*	*Disagree*	*Strongly disagree*
1	When I face a problem in learning grammar, I can take effective action to solve the problem.						
2	If I feel stressed about grammar learning, I know how to manage my feelings.						
3	Mindful of the importance of environment on grammar learning, I can devise methods for exerting control over the learning environment.						
4	When I try to learn grammar, I tend to procrastinate.						
5	If I feel satiated while learning grammar, I can consciously manage this feeling.						
6	In learning grammar, I always set clear goals and constantly assess my progress towards achieving my goals.						
7	I am satisfied with the effectiveness of my personalized methods to cope with the stressful grammar learning conditions.						
8	While studying grammar, I am responsible for the elimination of negative environmental factors and for reinforcing the positive ones.						

Item No.		*Strongly agree*	*Agree*	*Slightly agree*	*Slightly*	*Disagree*	*Strongly disagree*
9	In grammar learning, I can adjust to new learning environments to optimize my learning.						
10	If I feel bored in the grammar learning process, I know how to restore and revive my interest.						
11	From the beginning to the end of grammar learning, I try insistently to achieve my goals.						
12	If I feel stressed at any stage of my grammar learning, I can immediately manage to overcome such a feeling.						
13	I can find or create optimal learning environments to study and learn grammar.						
14	I consciously monitor my progress in learning grammar.						
15	My personal methods for minimizing my boredom in grammar learning are effective.						
16	I believe in my ability to accomplish my grammar learning goals during a specified period of time.						
17	When I come across difficult grammatical points, I easily get disappointed and give up learning.						

Item No.		*Strongly agree*	*Agree*	*Slightly agree*	*Slightly*	*Disagree*	*Strongly disagree*
18	Given my character and my preferred learning styles, I can identify distractions in the learning environment and act to reduce or eliminate them.						
19	I easily get distracted while learning grammar.						
20	I am confident about my ability to overcome satiation while learning grammar.						
21	I believe I am capable of attaining my grammar learning goals according to my learning schedule.						
22	I can manage my feelings to make grammar learning more enjoyable and effective.						
25	If I find the learning environment uncomfortable while learning grammar, I can identify and fix the problems.						
24	While learning grammar, I can effectively overcome procrastination tendencies.						
25	Given the time-consuming and tediousness of grammar learning, I have my own effective methods to stay on track and keep on learning grammar.						