

Assessing Learner Autonomy:
Development and validation of a localised scale

Lilan Lin *Anhui Jianzhu University, China*
Hayo Reinders *Unitec, New Zealand*

Abstract: With the inclusion of learner autonomy as an educational objective in many educational programs around the world, it is becoming increasingly important to be able to measure to what extent different interventions are successful. Although some general tests exist, we argue in this article for the need to develop localised instruments that are culturally appropriate, educationally relevant, and that are based on pedagogical principles shared by the staff and students in the particular context. Specifically, we describe the creation of a scale for measuring learner autonomy at a University in China. We detail the process of its development, from initial conception through to rigorous psychometric validation. In doing so, we will share some of the lessons learned and discuss some of the implications for scale developers and educational practitioners in other contexts.

Keywords: *learner autonomy, scale development, validity, reliability*

1. Introduction

Higher education in China has seen major changes in recent years. One of these changes relates to an increased focus on developing in learners a capacity for lifelong learning. This focus is also evident in the area of language education. A wide range of initiatives has been carried out across the country to encourage learners to take greater responsibility for their own learning. These include the establishment of self-access centres, the provision of online resources to complement classroom-based education, and more broadly a shift of focus in formal education to developing self-directed learning skills as well as a range of 21st-century learning skills, including critical thinking, collaboration, negotiation, and so on. In the area of language education, these and related concepts have usually been viewed through the lens of learner autonomy. The field of language learner autonomy has seen a great deal of academic activity in the last 30 to 40 years. There is now a large body of research and practitioners' experiences from an increasingly wide range of settings, both geographic and with learners of different backgrounds and ages, to enable educational providers to draw on as a foundation for the development of programs and resources to meet changing requirements within the Chinese education system. Although the concept of autonomy has been defined, operationalised, and implemented in many different ways, a shared aim of all such endeavours is to help learners develop the necessary attitudes, metacognitive awareness, as well as practical skills to take on an active role in determining their own learning.

In the Chinese context, many efforts have been made in recent years to draw on the existing research in order to implement a focus on autonomy in language education, primarily in the tertiary sector. It is important that the impact of such initiatives be investigated. However, as in other countries, this is hampered by the fact that autonomy is notoriously difficult to measure. As a concept that includes a psychological, as well as cognitive and a political dimension, the development of identifying changes in learners as a result of an educational intervention is a major challenge. Efforts in China so far have not been able to achieve this, largely because of a number of limitations in the available instruments. In this article we therefore propose a new scale to measure learner autonomy, to be used in Chinese institutions of higher education. We will begin by providing a brief review of learner autonomy and its assessment, before describing the Chinese context. In the main part of the article we describe the development of our scale for learner autonomy.

2. Learner autonomy and its assessment

2.1 Learner autonomy in China

Research on learner autonomy in the Chinese EFL (English as a foreign language) context began in the 1990s. The early research was mainly restricted to the introduction of the theories and practices of autonomy in western countries, and few empirical studies were carried out. Since the year 2004, however, there has been a remarkable growth of interest in the research on autonomy. There are specific reasons why the issue of learner autonomy is of great concern in the current Chinese tertiary EFL context. In a dramatic move to popularize higher education, China launched a program to boost university enrollment in 1999, which brought about a challenge to the way how *College English*, a compulsory course for all non-English majors in their first two years of university study across China, was taught. In addition to general criticism of the course and its perceived ineffectiveness and inefficiency, it was argued that *College English* should teach students how to use English effectively and enable them to exploit linguistic and other resources

available to them for their own purposes.

This led to the launch by the Chinese government of the ‘*College English Teaching Reform*’ and the publication of the ‘*College English Curriculum Requirements (Requirements hereafter)*’ in 2004. The *Requirements*, a set of official guidelines for College English teaching, states, for the first time in the history of such guidelines in China, that one of the objectives of *College English* is “to enhance students’ ability to study independently” (*The Requirements*, 2004, p.5) and it further emphasizes that “an important indicator of the successful reform of the teaching model is the development of individualized study methods and the autonomous learning ability on the part of students” (*the Requirements*, 2004, p.23).

Concurrent with the implementation of autonomy-based pedagogical programs in Chinese colleges and universities, the past decade has witnessed a proliferation of empirical research, conducted in six principal areas: 1) survey on students’ autonomous learning ability (Xu, et.al., 2004a); 2) effects of pedagogical programs on the development of autonomy (Xu, et.al., 2010); 3) relationship between strategy use and autonomy (Tan & Zhang, 2015); 4) effects of strategy training on autonomy (Shi, 2010); 5) use of technology on autonomy (Lv, 2015), and 6) relationship between autonomy and proficiency gains (Lin & Chen, 2009). The most commonly used method of eliciting data in the autonomy-related empirical research is the survey questionnaire. Unfortunately, scant attention has been paid to how assessment measures should be developed and used to ensure valid data analysis and, therefore, the quality of research. Built on a well-established framework, this paper reports on the process and methods used to develop a psychometrically-tested scale that can be used to assess Chinese tertiary EFL students’ autonomy.

2.2 Assessing autonomy

The assessment of autonomy is difficult as “autonomy is a multidimensional construct” (Benson, 2005: 51) and is “not a single, easily described behavior” (Little, 1991: 7). But this does not mean that “we should not attempt to measure it. If we aim to help learners to become autonomous, we should at least have some way of judging whether we are successful or not” (Benson, 2005: 54). Most research in autonomy is descriptive or qualitative. Self-assessments, portfolios and interviews are common instruments to elicit data from learners or their teachers about their perceived level of autonomy. Despite their usefulness in providing a deep understanding of issues in specific contexts, quantitative data would enable this to be complemented with broader generalisations, more objective judgment and significant labour- and time-saving benefits.

Attempts have been made to develop quantitative measurements of autonomy in language learning. Cotterall (1995:195) drew on data of learners’ beliefs to identify six factors underlying autonomy: the role of the teacher, the role of feedback, learner independence, learner confidence in their study ability, experience of language learning, and approach to studying. But she had difficulties in replicating the clusters of beliefs on which these factors were based in subsequent studies (Benson, 2005:79).

Targeting Hong Kong university students, Chan (2001) conceptualized autonomy as beliefs on the five dimensions: aims and motivation of language learning, perceptions of the teacher's’ role and their own, learning styles and preferences, learners’ perceptions of autonomous learning, and the extent to which they consider themselves as autonomous learners. Chan used a combination of interview questions about learners’ learning experience, their perceptions of learner autonomy and their own autonomy as well as three sets of questionnaires that asked about learners’ views on the teacher’s role, their own role and their learning styles. But she did not provide any psychometric data on the test and the development of the questionnaires.

Dixon’s (2011) 50-item measurement of autonomy was developed out of a long list of 256 items. Through exploratory factor analysis (EFA), he found six clusters: linguistic confidence, information literacy, social comparison, locus of control, metacognition and self-reliance. Dixon compared teacher estimates and questionnaire data to check its reliability and validity. Apart from the small scale data, Dixon’s measurement has been challenged for his way of “labelling of the factors” (Cooker, 2012:77). But Dixon concluded with an important message that learner autonomy may not be measured in an abstract sense, but using a questionnaire serves a useful purpose when the data it provides are viewed in context and in consultation with learner (2011:313).

Murase’s (2015) study involved two stages. Based on a conceptual model of autonomy within the technical, psychological, political-philosophic and social-cultural categories, Murase developed a 113-item scale out of the initial 143 items by conducting the confirmatory factor analysis (CFA) to 90 students, a small sample size for conducting factor analysis. According to Byrne (2001, cited in Worthington and Whitaker, 2006: 815), “the application of CFA procedures to assessment instruments that are still in initial stages of development represents a serious misuse of this analytic strategy”. In Stage 2, Murase conducted EFA and CFA among a group of 1517

students. But there is a lack of clarity in the procedures with respect to the use of samples in EFA and CFA, factorability, extraction methods and rotation methods.

The Learner Autonomy Profile (LAP) is a scale designed for general education rather than for language learning, consisting of four sub-scales. Following Confessore's (1981) conceptual model of autonomy, Meyer (2001), Carr (1999), Derrick (2001), and Ponton (1999), respectively, developed and validated instruments that assess learners' dimensions of "desire," "resourcefulness," "initiative," and "persistence" (cited in Confessore & Park, 2004). But the four instruments treat learner autonomy as an amalgam of psychological constructs of behavioral intentions, non-observable behaviors.

Apart from the autonomy measurements mentioned above, attempts have been made to gauge students' autonomy using scales designed for self-direction, such as the Self-directed Learning Readiness Scale (SDLRS) developed by Guglielmino (1977), the Oddi Continuing Learning Inventory (OCLI) developed by Oddi (1986), or the General Causality Orientations Scale (GCOS) designed by Deci and Ryan (1985) to understand the nature of causality of behavior.

As can be seen from the above, the challenges to the above measurements lie in both the operationalization of learner autonomy and the validation of the measurements. Since "autonomy is both contextually-variable and a matter of degree" (Benson, 2007: 23), "the culturally-constructed nature of the classroom needs to be taken into account" (Ho & Crook, 1995:237) in the development of autonomy-related measurement. Thus, measurement designed to assess Chinese students' autonomy needs to be appropriately contextualized to fit into its context and to match different aspects of autonomy with the characteristics and needs of its learners.

2.3 Assessing learner autonomy in China

Despite the growth of empirical research on autonomy in the Chinese EFL context, there are only three published studies on the assessment of EFL students' autonomy in the Chinese academic journals. For the majority of those that do use questionnaire survey (Fu & Yang, 2007; Lv, 2016; Ouyang & Xie, 2009; Shao & Zhao, 2011; Shi, 2010; Xiao, Wang & Cao, 2011; Zhang, 2009), however, there is a lack of transparency not only in how the construct of autonomy was conceptualized and operationalized, but also in how the instruments were developed.

Xu & Wu (2004b) designed a five-dimensional scale for English autonomous learning: 1) understanding teaching goals and requirements; 2) setting learning goals and plans; 3) effectively using learning strategies; 4) monitoring strategy use, and 5) evaluating the English learning process. They used the following procedures in the development of the scale: 1) Consultation with experts on the model of Chinese non-English majors' autonomy; 2) Contributions from colleagues to the item pool; 3) A pilot study. Xu & Wu's (2004) measurement focuses on self-management skills in the English learning process. But their research only used descriptive analysis and correlation analysis, which are not enough to support the validity of a newly developed scale.

Hu (2011) reported how he designed a two-dimensional scale to evaluate university students' autonomy in language learning. Based on Littlewood's (1996) model and using EFA, he developed a scale which consisted of willingness (motivation and confidence) and capacity (knowledge and skills in self-management).

Ren & Gao's (2012) scale focused on Oxford's (2003) model of autonomy: technical, psychological, social-cultural and political-critical autonomy. But the problem of the scale lies in the operationalizations of the four perspectives. For example, Ren & Gao operationalized technical autonomy simply as memory, cognitive and compensation strategies, psychological autonomy as changes in beliefs, affectives, metacognitive strategies and learning styles, while social-cultural autonomy was operationalized as social strategies. However, the assumption that successful strategy use is equivalent to learner autonomy is "not warranted by the best strategy research" (Little, 2000:23). Problems still exist in the political-critical subscale. Ren & Gao (2012) define political-critical as control of ideology, power and authority. No information was given as to why they defined it this way.

The findings from these studies indicate the need for further research in autonomy assessment in China. First, previous studies fail to present a comprehensive picture regarding the multi-dimensional nature of the construct of learner autonomy. As a result, there are some inherent limitations in the design of the scale to assess autonomy. The items in Xu & Wu's (2004) scale refer to self-management skills, those in Ren & Gao's (2012) to learning strategies. Although Hu (2011) included psychological autonomy in his study by adding "willingness" in his scale, the core of the psychological dimension—students' beliefs and consciousness about how they should take responsibility and how they see their ability to learn is not mentioned in the scale.

Second, development of the scales has not been accompanied by standardized psychometric steps, leading to possible concerns about their validity. Xu & Wu (2004) did not conduct a factor analysis to support the validity of the newly developed scale. Both Hu (2011) and Ren & Gao

(2012) only conducted EFA, but not CFA or validity analysis. The absence of a valid assessment measure makes it difficult to interpret research findings and make generalizations.

As assessment is “essentially social activities, influenced by unique affordances and constraints of a particular educational context” (Matuga, 2006:317), it is clear, from the above, that a significant amount of work remains to be done in China, as well as in other contexts, to develop meaningful, locally-relevant tools for measuring autonomy. Our purpose in the rest of this article is to describe, in some detail, the process that we went through in constructing a scale to be implemented in one university in eastern China.

3. Methods

3.1 The context

This research was conducted in a Chinese university, where *College English* teaching, as in most colleges and universities in China, was characterised by large class sizes and a teacher-centered approach. The impetus for the curriculum reform in this university came from dissatisfaction with the status quo in teaching and learning on the part of the teachers, the students, as well as the university authorities. Dissatisfaction existed in particular with regards to a lack of individualisation, meaning that learners were expected to proceed at the same pace, a lack of freedom of choice for students, lack of concern about individual differences, the widespread use of summative assessment, and an extreme dependence on the teacher. In response to the *Requirements*, a teaching reform was initiated at the university, and funded by the local government, to promote learner autonomy through innovation of the curriculum and the teaching modes employed.

It was decided that a major focus of the reform should be identifying ways in which students can be encouraged and supported in taking more control of their learning. An extensive engagement with the literature and prior operationalizations of learner autonomy led to our conceiving the construct as consisting of three main elements (Candy, 1991; Holec, 1981; Little, 1991; Reinders, 2011; Wenden, 1991): 1) learner’s ability in self-management of their language learning; 2) learner’s consciousness of and attitudes towards autonomy; and 3) autonomous learning practice. The current research focused on all three aspects and empirically tested this hypothetical construct.

3.2. Participants

Sophomores were chosen to complete the questionnaire anonymously. The reason we decided to choose them was that they had had one-year self-access listening experience during the time surveyed. It was assumed that they could give more information about this mode based on their experiences. The present study used three samples. Sample 1 provided the data for the EFA, Sample 2 for the CFA and Sample 3 for a second CFA. One of the criteria for data in factor analysis is the sample size. The proposed guideline for minimum ratios of participants to items is 5:1 or 10:1 (Worthington & Whittaker, 2006: 817). As there were 52 items in the initial version of the scale, Sample 1 comprised 588 subjects. Sample 2 comprised 380 subjects for the 42-item scale and Sample 3 was another 455 subjects for the 38-item scale.

3.3. Data analysis procedures

The development of the scale uses a process recommended by Worthington and Whitaker (2006). Following an extensive initial conceptual development stage, item creation was undertaken and tested using the Delphi technique, EFA and CFA, before examination of its reliability and validity.

All quantitative survey responses were coded, entered into the data editor and analyzed using the SPSS 10.0 in the EFA and reliability analysis, and AMOS (analysis of moment structure) version 6.0 in the CFA.

4. The process of constructing a scale for learner autonomy

4.1 Item generation

The first task was the creation of the item pool to reflect the content of the target construct, that is, the three dimensions of learners’ attributes in autonomy. Self-management skills were demonstrated as the following skills students believe that they possess: defining objectives, defining and selecting appropriate materials and activities, defining pace of learning, monitoring learning, evaluating the process and evaluating the outcomes (Holec, 1981). Learners’ psychological autonomy focused on their consciousness and willingness to take control of one’s learning as reflected in their attitudes towards their own role, their confidence to take control of their own learning (Little, 1991; Reinders, 2011; Wenden, 1991), and their meta-cognitive and meta-linguistic awareness interpreted as their knowledge about how well they perform on learning tasks, and the demands and procedures of second language learning, and strategies for English learning (Broady, 1996; Reinders, 2011). Learning practice was interpreted by a series of specific

behaviors in natural context of learning that show how learners exercise their self-management skills and their attitudes and awareness towards autonomy (Candy,1991; Hedge, 2000; Reinders, 2011).

The draft scale was composed of three parts. The first part was a brief demographic questionnaire for the purpose of collecting information about students' name, class, and their previous English learning experience. In the second part, students self-reported (using a five-point Likert scale) their learning management skills, their attitudes and consciousness of autonomy on 44 questions. In order to make a comparison between what students think they are capable of and/or feel what they should do, and what they actually do, an open-ended questionnaire was designed in order to understand what kinds of learning practice were regarded as autonomous by students. Eighty juniors at the University who had been using self-access listening resources in the past two years and 10 teachers of English at the University were asked to list 20 activities associated with autonomous learning behaviors. Their lists of items were collected and compared with what the literature describes about the profile of an autonomous language learner and then summarized into 28 items, which could be thought of as manifestations of autonomous language learning behaviors. These 28 items formed the initial third part of the questionnaire, again using a five point Likert scale.

4.2 Validation

To test content validity of the scale the Delphi technique (Linstone & Turoff, 1975) was used to gain consensus among an expert panel on the content validity of the 72 items thought to reflect learner attributes in autonomy. The expert panel comprised 11 expert teachers from four different universities. Each panel member was asked to complete the questionnaire anonymously and independently, and evaluated each item to determine the degree to which the items measured the three constructs—self-management skills, autonomous psychology and autonomous behaviors among Chinese tertiary EFL learners. After two rounds, and having excluded items for which less than 80% agreement was reached, 52 items were retained from the initial 72 item pool.

In order to test construct validity a combination of exploratory and confirmatory procedures was followed to confirm the factorial stability of the proposed scale. According to Fabrigar *et al.* (1999) EFA is used in the initial stages of scale development to explore potential hypothetical relationships between factors and items, with CFA used subsequently to validate such relationships. Before the EFA, an item analysis was performed through corrected item-total correlation, which was used for purification purposes because unimportant items may confound the interpretation of the factor analysis. Eight items were discarded because they had low (less than .30) corrected item-total correlation, which improved the reliability of the scale from Cronbach's Alpha .932 to .952 .

Then the suitability of the data for EFA was assessed by means of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. For this research, the KMO computed was 0.956 (>0.7) and Bartlett's test correlation matrix was .000 (Table 1). The results show that the sample was appropriate for exploratory factor analysis.

Insert Table 1 here **Table 1 KMO and Bartlett's Test**

EFA assesses the construct validity and determines how many latent variables underlie the complete set of items during the initial scale development so that the items that do not measure an intended factor may be eliminated. In the EFA phase, first, factor extraction using principal component was conducted to determine the smallest number of factors that could be used to best represent the interrelations among the set of variables. According to Kaiser's criterion, only factors with an eigenvalue of 1.0 or more are retained for further investigation. The resulting eigenvalues were plotted, revealing 3 distinct factors arising from the item pool to be retained for interpretation (Lin, 2013, Table 2).

Insert Table 2 here **Table 2 Exploratory factor analysis**

Second, factor rotation was employed in order to present the pattern of loadings in a manner that is easier to interpret. According to Comrey and Lee's (1992) pattern coefficients greater than .71 are considered excellent, greater than .63 are very good, greater than .55 are good, greater than .45 are fair and greater than .32 are poor. Two items did not load on any of the components using a cutoff loading of .45 and were subsequently eliminated from the scale. After poor items were deleted from the scale, a 42-item scale was arrived at (Table 2).

Factor 1 was clearly the most important one since it accounted for 33.6% of the total scale variance (Table 2). It was defined by 16 of the pool items. Since all these items had to do with the students' use of self-management skills, it was labeled Self-management Skills.

The statements clustered as Factor 2 were defined by another 16 items and accounted for 12.6% of the variance (Table 2). Since this factor relates to students' autonomous learning behaviors such as planning, implementing, monitoring and evaluating learning activities, and previewing and reviewing strategies and strategies to improve their learning outcomes, this factor was labeled Autonomous Learning Behaviors.

Factor 3 consisted of 10 items and accounted for 4.2% of the variance (Table 2). An examination of this factor revealed that all these items represent students' willingness and consciousness to take responsibility for their own learning (Item 2, 7, 8, 9 and 10), their confidence in their ability to work independently of the teacher (Item 17 and 20) and their meta-cognitive and meta-linguistic awareness in language learning (Item 21, 23 and 25). This factor therefore was named Autonomous Psychology.

CFA is intended to test the hypothesis that a relationship between the observed variables and the underlying latent construct exists and to confirm the extent to which the new scale is replicated in the new sample data. To continue the investigation of the construct validity of the developed instrument produced by the above EFA, CFA was conducted in a separate sample using the developed 42-item version of the scale from the EFA. The CFA was specified using the model extracted from the EFA. The structural equation modeling (SEM) confirmatory procedure was taken using AMOS to find good fit of the model to the data in the new samples and support the factor structure of the scale. Since no non-significant paths were found for possible item deletion in model specification and estimation phases, model fit was evaluated by a suggested minimum collection of the fit indices by Kline (2005, cited in Worthington & Whittaker, 2006:829), which included: 1) the ratio of Chi-square to degrees of freedom, which shows how well the model fits the data with ratio of χ^2 to $df \leq 2$ or 3 indicating acceptable fit; 2) a value of Root mean squared error of approximation (RMSEA) of about .06 or less as acceptable model; and 3) the Comparative Fit Index (CFI), and Tucker-Lewis index (TLI) with the value of .90 indicating acceptable model fit.

The indicators of goodness-of-fit obtained through the analysis of the 42-item model did not indicate a good fit using the above cut-off line (TLI= .847; CFI= .855). Modifications were sought four rounds based on the modification index (M.I) (Table 3). One common strategy is simply to remove the parameter with the largest M.I, then see the effect as measured by the chi-square fit index. But modifications should be done on the basis of theory, not just the magnitude of the M.I. This process yielded an improved fit over the previous models ($\chi^2/df=2.283$, $p < .001$; TIL=.882; CFI = .889; RMSR =.058) and has 38 items altogether. In order to determine the stability to establish predictive validity of the revised model, Sample 3 was used and the re-specified model tested in a second confirmatory approach, which shows that the three-factor model for the 38-item AELS fitted the Sample 3 data well ($\chi^2/df=2.487$, $p < .001$; TIL=.912; CFI = .917; RMSR =.057). (Lin, 2013)

Insert Table 3 here **Table 3 Modification Indices of M1, M2, M3, and M4**

Once the final model was established from the CFA, the internal reliability of the modified AELS was evaluated.

3.5 Reliability

The measures of test reliability make it possible to estimate as to what proportion of the total item score is error variance. An internal consistency, which refers to the degree to which the items measure the same underlying construct, was evaluated. One of the most commonly used indicators of internal consistency is Cronbach's alpha coefficient. Ideally, Cronbach's alpha coefficient should be over .7 before we can say that the scale is internally consistent. As can be seen from Table 5 the computed values of Cronbach's coefficient alpha for the global scale, and the three subscales were .949, .934, .899, and .910 respectively, which is considered to have an excellent level of internal consistency of the items in the global scale as well as in the three subscales.

Another measure of the internal consistency for this research was the split half test. We split the test with the 380 sample into two and calculated the respondents' scores on each of the half test. The split half tests are strongly correlated if the correlation coefficient is over .8. For this research, the split-half coefficient of the global scale was .919 while the split half coefficients for the three subscales varied between .902 and .838.

3.5 Convergent and discriminant validity

To determine the construct validity of the AELS, both convergent validity and discriminant validity were assessed in the students from Sample 3. Convergent validity is the degree that a trait is well measured by its indicators while discriminant validity is the degree to which measures of different traits are unrelated (Campbell & Fiske, 1959). In this study, convergent validity was

assessed by factor loading, the Average Variance Extracted (AVE) and Composite Reliability (CR).

Table 4 shows the standardized factor loading of all the indicators in each factor is more than acceptable level (more than .5) after deleting six low factor loading indicators. According to the criterion of Fornell & Larcker (1981), all the three factors have a high level of AVE between .702 and .737 while their values of CR are .872, .935 and .934, indicating high internal reliability.

Insert Table 4 here **Table 4 Convergent validity**

For discriminant evidence, the cross loadings between indicators were examined. As shown in Table 5, indicators load much higher on the factor they measure than on the other factors, indicating good discriminant validity of the final 32-item AELS.

Insert Table 5 here **Table 5 Matrix of loadings and cross-loadings**

To sum up, the above tests for the final model showed that the 32 items consistently measure the same construct of the AELS and the items within each factor were highly correlated with each other.

5. Conclusions

It is clear from this article that autonomy has received a considerable amount of interest in China in recent years. A review of existing studies and in particular reports on measurements of autonomy, however, show that there are number of significant issues that will need to be addressed in the future. The first of these, and foundational to all the others, is that there is a lack of clarity around the concepts in relation to the topic of autonomy. Without a shared understanding of what autonomy entails, it will be difficult to move the field along. At the very least, it is important for authors to clarify the theoretical framework on the basis on which they investigate autonomy.

Another finding was that existing instruments do not adequately report on the psychometric aspects of their development. It is important that future reports include all relevant information, so that a clear assessment can be made of the value of any instrumentation used.

As a result of the above there is still a significant need for work to be done in order to create reliable instruments that have high construct validity and high relevance in specific contexts. As the focus in education in China shifts increasingly on developing learners' ability to manage their own learning, both within and outside of formal education, this becomes all the more important.

References

- Benson, P. (2005). *Teaching and Researching Autonomy in Language Learning*, Beijing: Foreign Language Teaching and Research Press.
- Benson, P. (2007). Autonomy in language teaching and learning. *State of the Art Article Language Teaching*, 40(1): 21-40.
- Broady, E. & Kenning, M. (1996). Learner autonomy: an introduction to the issues. In: Broady, E. & M. Kenning. *Promoting Learner Autonomy in University Language Teaching*. London: CILT, p.9-21.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2):81-105.
- Candy, P. C. (1991). *Self-direction for Lifelong Learning*. San Francisco, US: Jossey-Bass.
- Chan, V. (2001). Readiness for learner autonomy: what do our learners tell us? *Teaching in Higher Education*, 6(4): 505-518.
- Comrey, A. L., & Lee, H. (1992). *A First Course in Factor Analysis*, New York: Academic Press.
- Confessore, G. J., & Park, E. (2004). Factor validation of the learner autonomy profile, version 3 and extraction of the short form, *International Journal of Self-directed Learning*, 1(1):39-58.
- Cooker, L. (2012). *Formative (self-) Assessment as Autonomous Language Learning*, (PhD). University of Nottingham.
- Deci, E. L. & Ryan, R. M. (1985). The general causality orientations scale: self-determination in personality, *Journal of Research in Personality*, 19:109-134.
- Guglielmino, L.M. (1977). *Development of the Self-directed Learning Readiness Scale*, (PhD). University of Georgia.
- Cotterall, S. (1995). Readiness for autonomy: investigating learner beliefs, *System*, 23(2): 195-205.
- Dixon, D. (2011). *Measuring Language Learner Autonomy in Tertiary-level Learner of English*, (PhD), University of Warwick.

- Fabrigar, L.R., MacCallum, R.C., Wegener, D.T., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research, *Psychological Methods*, 4(3):272-299
- Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1):39-50.
- Fu, L. F. & Yang, J. D. (2007). Research of students' autonomous English learning abilities in multimedia and web-based learning environment, *Foreign Languages and their Teaching*, 223(10): 36-38.
- Hedge, T. (2000). *Teaching and Learning in the Language Classroom*, Oxford: Oxford University Press.
- Ho, J., & Crookall, D. (1995). Breaking with Chinese cultural traditions: learner autonomy in English language teaching, *System*, 32(2): 235-244.
- Holec, H. (1981). *Autonomy in Foreign Language Learning*, Oxford: Pergamon.
- Hu, J. H. (2011). Assessment of autonomous foreign language learning capacities: designing a two-dimensional questionnaire, *Foreign Language World*, 145(4): 12-17.
- Lin, L. L. (2013). Development and validation of the Autonomous Learning Scale. *Foreign Language World*, 157(4), 73-80.
- Lin, L. L., & Chen, Y. E. (2009). An investigation into psychology, learning behaviors and outcomes of tertiary students in resourced-based learning environment, *Journal of PLA University of Foreign Languages*, 32(2): 45-50.
- Linstone, H. A., & Turoff, M. (1975). *The Delphi method: Techniques and Applications*. Boston, MA: Addison-Wesley.
- Little, D. (1991). *Learner Autonomy. 1: Definition, Issues and Problems*, Dublin: Authentik.
- Little, D. (2000). Strategies, Counseling and Cultural Difference: Why we Need an Anthropological Understanding of Learner Autonomy, In: Ribé, R. *Developing Learner Autonomy in Foreign Language Learning*, Barcelona: University of Barcelona, p. 17-33.
- Lv, T. T. (2015). A study on the flipped classroom based college English autonomous learning mode, *Foreign Languages in China*, 69(1): 77-83.
- Ministry of Education of the People's Republic of China. (2004). *College English Curriculum Requirements (For Trial Implementation)*. Beijing: Qinghua University Press.
- Murase, F. (2015). Measuring language learner autonomy: problems and possibilities. In: Everrhard, C. J., & Murphy, L. *Assessment and Autonomy in Language Learning*, NY: Palgrave, p. 35-63.
- Matuga, J. M. (2006). The role of assessment and evaluation in context: pedagogic alignment, constraints and affordances in online courses. In: Williams, D.D., Howell, S., & Hricko, M. *Online Assessment, Measurement and Evaluation: Emerging Practices*. London: Information science publishing, p.316-330.
- Oddi, L. (1986). Development and validation of an instrument to identify self-directed continuing learners, *Adult Education Quarterly*, 36(2): 97-107.
- Ouyang, J. P., & Xie, J. (2009). Constructing a model of autonomous English learning: an empirical study, *Journal of PLA University of Foreign Languages*, 32(3): 46-51.
- Reinders, H. (2011). Towards an operationalisation of autonomy. In: Ahmed, A. Cane, G. & Hanzala, M. *Teaching English in Multilingual Contexts: Current Challenges, Future Directions*. Cambridge: Cambridge Scholars Publishing, p. 37-52.
- Ren, H. H. & Gao, P. (2012). Development of English autonomous learning skill rating scale for college students, *Journal of University of Science and Technology Liaoning*, 35(2): 210-214.
- Shao, S. Y. & Zhao, R. (2011). Metacognitive strategy and autonomous learning: an investigation into self-access extensive reading, *Computer-assisted Foreign Language Education*, 141(5):60-64.
- Shi, J. F. (2010). Fostering science-and-technology students autonomy in web-based learning environment: an empirical study based on metacognitive-strategy training, *Foreign Languages in China*, 34(3):63-67.
- Tan, X. & Zhang, Z. H. (2015). Relationship among English learning strategies, learner autonomy and CET-4 scores, *Foreign Language Learning Theory and Practice*, 149(2):59-65.
- Xiao, W. Y., Wang, X. P., & Cao, Q. Y. (2011). An empirical study on metacognitive strategy training used in improving learner autonomy and learning achievement, *Foreign Language Research*, 159(2): 109-113.
- Xu, J. F., Peng. R. Z., & Wu, W. P. (2004a). Study of non-English majors' autonomous learning ability, *Foreign Language Teaching and Research*, 36 (1):64-68.
- Xu, J. F., & Wu, W. P. (2004b). Fuzz comprehensive measurement of students' autonomous learning ability, *Research in Higher Education of Engineering*, 3:84-86.
- Xu, J. F., Tang, F., & Liu, Z. H. (2010). A three-in-one teaching model to promote freshmen's autonomous learning, *Foreign Language Education*, 31(6): 60-64.
- Wenden, A. (1991). *Learner Strategies for Learner Autonomy*. London: Prentice Hall

International.

- Worthington, R. L., & Whittaker, T. A. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices, *The Counseling Psychologist*, 34 (6): 806-838.
- Zhang, L. (2009). Metacognitive strategy training: a prerequisite to fostering autonomous learning ability, *Foreign Language World*, 133(4):56-60.