Iranian TEFL Graduates' Conceptions of Measurement Error in Research: A Genealogical Narrative Inquiry

Azar Hosseini Fatemi

Ferdowsi University of Mashhad, Iran

hfatemi@ferdowsi.um.ac.ir

Saeedeh Shamsaee

Ferdowsi University of Mashhad, Iran

sa_sh756@stu-mail.um.ac.ir

Mohammad Ali Shams

Ferdowsi University of Mashhad, Iran

mo.shams@stu-mail.um.ac.ir

Abstract

The aim of this study is to investigate Iranian TEFL graduates' conception of measurement error in research. Adopting a sequential explanatory multi-method strategy (Borg, 2009), the researchers analyzed causal and temporal relations in the research narratives elicited from 30 TEFL graduates. Gee's (1986) framework for identifying narrative discourse units (lines, stanzas, and episodes) was adopted to investigate participants' conceptions of logical orders in measure development algorithms and their knowledge of error sources. In addition, taking a narrative positivistic approach, the narratives were rated based on Optimal Matching Analysis (OMA). Finally in 'continuous event history modeling' phase of the study, Cox Proportional Regression Analysis showed how temporal markers in research narratives can be used to predict one's knowledge of measure development in research design. The results suggest that researchers' error-awareness and algorithmic knowledge correlate significantly with each other and constitute knowledge of measure development in general. The contribution of dimensionality and validity testing to this knowledge was also found to be statistically significant.

1. Introduction

Accurate measurement is central to scientific research. A measure of a construct is not the construct itself, but one of the various error-filled ways of measuring it (Viswanathan 2005). These are basic notions upon which psychometric theories and statistical procedures are based. Standards for measure development process in human sciences in general, and TEFL in particular, have been defined and frequently refined during the past 5 decades.

Quite naturally, it is expected that researchers who are to develop a measure for investigating the variables involved in teaching of English as a foreign language consider and apply the standards in error measurement to minimize the degree of error, and consequently, to increase the validity of the obtained results, and also to expand the generalizability of their findings. As Viswanathan (2005) maintains, there is no substitute for well-designed, reliable, valid measures. In this regard, measures developed within the M. A. final theses for TEFL in Iran are no exception. Taking a diagnostic approach, the present study aims to investigate how and to what extent the standard procedures for measure development are being regarded by Iranian TEFL researchers to achieve more dependable results.

2. Review of the Related Literature

The materials presented in this section are threefold. First, two major factors constituting researchers' knowledge of measurement error namely algorithmic knowledge of measure development steps, and knowledge of error sources are defined based on the literature. Second, possibilities for using qualitative and quantitative narrative inquiries in TEFL research are introduced. And finally,

the methodology and findings of two recent works in which TEFL practitioners' conception of research have been studied are discussed.

2.1. Knowledge of Measure Development

Knowledge of measure development can be operationally defined as the combination of two major lines of understanding namely that of algorithmic knowledge and knowledge of error sources (Viswanathan, 2005). According to Viswanathan (2005), the measure development process consists of a series of steps to develop reliable and valid measures. It starts with carefully understanding what is being measured, through definition of the construct and delineation of its domain. After generating the items, internal consistency reliability, test-retest reliability, dimensionality, and validity tests have to be performed on the designed measure with items being added, modified, or deleted along the way.

2.2. Taxonomy of Measurement Error Sources

Past research has categorized sources of errors in several ways. Bardo, Yeager, and Klingsporn (1982) differentiate between respondent-related errors that are content-specific and format-related errors that are due to respondents' use of response formats. Method variance has been described as varying in level of abstraction from item wording and scale types to halo effects and social desirability (Bagozzi and Yi 1991). Ghiselli, Campbell, and Zedeck (1981) distinguish between situation-centered and person-centered sources of errors across administrations. Viswanathan (2005) has developed a rather inclusive taxonomy of measurement error sources. In this taxonomy, all the sources of errors either arise of certain respondent characteristics, or certain

characteristics of the method, or interactions between these two categories. He calls the former individual-related sources (idiosyncratic and generic), and the latter method-related sources of error (item content, response format, and administration issues).

2.3. Quantitative Approaches to Narrative Analysis

An interest in narrative and the use of textual evidence does not prevent the use of quantitative techniques. Whereas the term 'qualitative approach' is often used to imply both an interest in qualitative data and the rejection of statistical methods of analysis, it is clear from the work of researchers such as Franzosi (1998), and Bearger and McAdams (1999) that the use of statistics is perfectly compatible with certain types of analysis of qualitative material. Although many scholars have stressed the rich nature of qualitative data in comparison with less detailed quantitative data, and qualitative data have been argued to take better account of contextual factors, largely ignored by quantitative studies, recent developments in quantitative narrative analysis have demonstrated that this conceptualization of distinction between qualitative and quantitative evidence is somewhat oversimplistic. Czarniawska (2004) discusses event history modeling, optimal matching analysis, and the construction of case histories as three main quantitative methods of analyzing narrative data.

In many respects event history modeling resembles more widely understood regression techniques, such as ordinary least squares (OLS) regression and logistic regression (where the dependent variable is dichotomous). However, event history modeling differs from Standard Multiple Regression in that the dependent variable is not a measurement of an individual attribute such as IQ, rater it is derived from the occurrence or non-occurrence of an event, which is temporally marked. The types of data that are

suitable for analysis using event history techniques can, therefore, be characterized as having a temporal dimension. Retrospective reports, relying on individuals' memories of the timing of events can be used to collect the data for event history modeling analysis. In essence, this technique allows us to evaluate the relative importance of a number of different categorical or continuous variables for predicting the chance of an event occurring. Generally there are two types of methods used to study the temporal dimension of events in event history modeling namely 'continuous time methods' and 'discrete time methods'. While the former assumes that the time at which an event occurred has been exactly measured, the latter is usually used when the unit of time is relatively large (years, or months perhaps).

Although event history techniques are powerful and flexible, they still have the disadvantage that they do not deal with sequences holistically. As Abbott (1992) has suggested, an alternative approach to the analysis of event history data is not to attempt to model the underlying processes, which result in particular narrative realizations within the observed data, but rather to focus on the narratives themselves and to try to establish a systematic description or typology of the most commonly occurring patterns within them. This approach has been termed 'narrative positivism'. Abbott and Tsay (2000) have demonstrated the use of optimal matching techniques to handle larger samples of sequence data. The basic concept behind optimal matching analysis (OMA) is that in order to be able to produce a typology of sequences it is necessary to be able to form clusters of similar sequences, and this is only possible if a measure of the difference between each pair of sequences can be derived. In order to calculate this measure of difference, OMA counts how many elementary operations (substitutions, insertions, or deletions) are needed to turn one sequence into another. This technique allows researchers to analyze

sequences of data holistically, instead of having to focus on specific transitions between states.

2.4. Oral Narratives as Genealogical Markers of Knowledge

The structure of the participants' narratives can reflect the elements and developmental processes involved in formation of their research projects. This structure is discernable through analyzing the transcripts of the oral narratives produced by the participants answering the interviewer. Transcribing practices can be understood as ranging from those that attempt to record every detail of the verbal interaction to those that aim to preserve only the words which are spoken. At one extreme then, some researchers choose not to record any of the extra verbal material captured on the research tape such as pause, intonation, etc., and in addition may remove repetition, false starts, and non-lexical utterances such as 'umms'. Providing a clean transcript focuses on the content of what was said. At the other end of the spectrum, when the focus is not only on the content of the narrative but the way that a narrative is recounted is also salient, it is important to record the delivery of speech more faithfully.

An approach to transcription which perhaps could be understood as lying somewhere between the two extremes has been used by a number of researchers with an interest in narrative elements of qualitative interviews. It is informed by the work of Gee, a sociolinguist who has worked on the structure of oral language. Gee (1986) suggests that despite the variation in discursive styles between individuals of different ages and from different cultural backgrounds, the basic units of discourse that organize its structure are maintained. Using Gee's framework the smallest unit of discourse is the 'line'. Each line is made up of a short sequence of words comprising one 'idea unit'.

The ends of lines are typically marked by the speaker with a short pause and a fall in the pitch of the voice. Within an oral performance these lines are typically grouped together to form stanzas. The lines within a stanza will often display a parallel structure so that they match each other in terms of content or topic. The breaks between stanzas are usually marked by a longer pause on the part of the speaker. The larger unit or the section is what we might think of as an episode within a longer narrative. The beginning of a new section is often marked by a good number of hesitations or false starts. By applying Gee's typology of the different units within a discourse it is possible to produce a transcript which preserves some of the rhythm and structure which characterizes speech.

2.5. English Language Teachers' Conception of Research

In educational research, narrative has emerged as both a method in, and an object of, inquiry in teacher education, specially in description and analyses of teacher knowledge (Connelly and Clandinin, 2006). Researchers of teacher education argue that teacher knowledge is largely structured through stories and that the story is epistemologically the most authentic way to understand teaching from the teachers' point of view. Using narrative inquiry to investigate the formation and organization of participants' knowledge in applied linguistics has recently gained more attention from TEFL researchers. The last decade has observed the growing number of studies on language teachers' conception of research (Nunan 1997; Allison and Carey 2007; Atay 2008; Borg 2006; 2007; 2008; 2009). Xu and Liu (2009) explore teachers' assessment knowledge and practice through a narrative inquiry of college EFL teachers. Drawing on Crites' (1971) notions of sacred stories and secret stories in teachers' professional knowledge landscape, they examine teachers' account of their

experience of assessment reform. Craig (2007) believes that through telling and retelling teachers' knowledge and practices in educational reforms can be interpreted and understood narratively. Xu and Liu (2009) use three sets of data. The first set includes two formal semi-structured interviews. The second set of data is derived from personal communications with the teachers. These discussions were used as input to formulate the interview questions. The third set of data includes teachers' teaching diaries. They discuss three structural conditions of teacher knowledge, temporality, sociality, and place (Connelly and Clandinin 2006), and their effects on teachers' knowledge construction of assessment.

An interest in teacher research engagement is evident in the literature on English language teaching, though in this field only a limited number of empirical studies of teachers' conceptions of research exist. McDonough and McDonough (1990) surveyed the views of research of teachers of English as foreign language, while Brown et al. (1992) report a survey of members of an international association for ELT professionals. These studies reported notions of research closely tied to quantitative and statistical methods. More recent work by Allison and Carrey (2007) reflects increasing empirical interest in English language teachers' engagement in research. Borg (2009) examines the conceptions of research held by teachers of English around the world. He argues that an understanding of this issue is central to the development of informed policies for promoting teacher research engagement, but relevant systematic evidence is lacking in the field English language teaching. This study shows that the teachers held conceptions of research aligned with conventional scientific notions of inquiry. The findings of this study point to a number of attitudinal, conceptual, procedural, and institutional barriers to teacher research engagement.

3. Research Questions

This study sets out to answer the following research questions:

1) To what extent the Iranian TEFL graduates are aware of the error sources in measure development process?

2) Is there any significant relationship between the Iranian TEFL graduates' knowledge of error sources and their knowledge of the logical orders in measure development algorithm?

3) What areas of Iranian TEFL graduates' knowledge of the logical orders in measure development algorithm contribute significantly to their general knowledge of measurement error in research?

4. Method

The method of the present study is based on what Borg (2009) calls a sequential explanatory multi-method strategy. This is a design which comprises collection and analysis of qualitative data followed by quantitative techniques.

Given the potentials of narrative analysis, the researchers have looked at the data to investigate the participants' conception of measure development in TEFL research design using both kinds of inquiry.

4.1. Participants

The participants of this study comprise 30 Iranian graduates of Teaching English as a Foreign Language from 5 state universities in Tehran. All the participants have had attended to independent measure development to investigate learning-related variables in their final thesis. Out of 58 cases which were firstly identified based on the mentioned criterion, 34 were accessible for

the researchers and accepted to participate in the study. Afterwards four of the research narratives were omitted from the sample because they lacked some of the main pieces of information required for narrative analysis; arguably it can be inferred from the collected data that the loss was mainly due to memory problems and very low narrative competence. To regard standards of ethics in research, all the informants were fully informed on the general aim of the study (of course avoiding revealing the unnecessary details that may have affected their performance), and were ensured about the confidentiality of the personal information they revealed.

4.2. Instrumentation

The main instrument used in this study was a semi-structured narrative interview during which participants told the story of the initiation, development, and accomplishment of their final thesis projects with the main focus being on the process of measure development. To prevent participants from telling digressing stories, general guidelines were provided by the interviewer to ensure that the order of events in the recounted narratives reflects that of actual situation as much as possible. In narrative research, interviews are usually preferred to questionnaires. Ochberg (1996) believes that questionnaires limit our informants to a narrow menu of preselected questions and answers, whereas interviews let informants choose the events that matter to them; the use of narrative within interviews clearly does give informants more opportunity to become more active subjects and to select what they believe to be the most salient information (Elliot, 2005).

4.3. Data Analysis

The sequential explanatory multi-method strategy determined the order of events in the present study. First the informants participated in a narrative interview session and provided the interviewer with personal narratives recounting the events in the accomplishment of their final M.A. thesis. The narratives were recorded digitally and saved as audio files on a computer. Then the stories were transcribed and analyzed both qualitatively and quantitatively.

4.3.1. Causality and Temporality in Narrative Discourse Units

In qualitative analysis, Gee's (1986) framework for identifying narrative discourse units was adopted to investigate both content and form of the narratives. While causal links and temporal junctures in the form of research narratives reflect the participants' degree of commitment to follow the standards for measure development, the content of the stories reveals the participants' sensitivity to various types of measurement error sources. After transcribing the narratives based on Gee's (1986) transcription model, units of narrative discourse namely lines, stanzas, and sections (narrative episodes) of the transcribed stories were indentified. Then lines signaling causal links and temporal junctures were marked and clustered into separate parts based on the association with defined steps in measure development process. In addition, all the references to possible sources of error were separately identified and listed for each participant.

4.3.2. Optimal Matching Analysis

In the first part of quantitative analysis, taking a narrative positivistic view, the researchers analyzed most commonly recurring patterns in designing and

developing research projects conducted by Iranian graduates of TEFL. To measure the extent to which procedures applied by the participants for doing their research match or mismatch the standard procedures suggested by research theorists, an Optimal Matching Analysis (OMA) was run. In order to calculate this measure of difference, OMA counts how many elementary operations (substitutions, insertions, or deletions) are needed to turn the sequence of events observed in participants' stories into the standard sequence of events as it is expected. In this way, each participant's conception of measure development was scored based on the extent to which it matched the standard algorithm (Viswanathan 2005).

4.3.3. Cox Proportional Regression Analysis

Cox Proportional Regression analysis is used to study the relationship between temporal variables. What makes this model more appropriate for analyzing the relationship between participants' knowledge of measure development algorithm is that it takes account only of the order in which events occur rather than the exact timing of events. There are 5 independent temporal variables in the model signaling the time spent for the accomplishment of domain delineation, item generation, reliability testing, dimensionality testing, and validity testing in measure development process respectively. The dependent variable is knowledge of measure development which has already been measured through OMA.

5. Results

The results of the study comprise both qualitative and quantitative findings are presented here in three subsections. First, researchers' conceptions of different

sources of error are discussed. Second, the relationship between researchers' algorithmic knowledge and their error-awareness is explored. And finally, the contribution of each of the steps in measure development process is analyzed through event history modeling.

5.1. Researchers' Conceptions of Error Sources

Analysis of the content of research stories told by the participants show that in general Iranian TEFL researchers are more sensitive to individual-related sources of error either idiosyncratic or generic, and that they are, most of the time, ignorant of method-related sources of error that may endanger the validity of the scales they have developed for measuring operational variables of their study. As it can be seen in Table 1, 76% and 45% of the participants were aware of possible idiosyncratic and generic individual-related sources of error that might have affected the final results and interpretation respectively.

Only a few participants thought that the specific context of administration may have affected the subjects' performance in their study. Among the 16% who mentioned possible item content as an error-yielding factor in their measurement, only one researcher reported any attempt to cover for this problem by changing the content of the items in the developed measure.

Among the method-related sources of error mentioned in the stories, response format and item content have the most and second most frequent types and tokens respectively. Among individual-related sources of error, idiosyncratic factors are more frequent, both in type and token, compared to generic factors.

The last column of the table shows the proportion of the number of type of the error sources that have actually been mentioned in the research narrative to those that are expected to be observed by them in the ideal situation and based

on the taxonomy proposed by Viswanathan (2005). The numbers shown here refer to collective proportional measures. The individual proportions have also been used to calculate error-awareness scores for each individual participant.

For instance, if a participant mentions half of the expected sources of error in item-content category, his score in that area will be 50 (out of 100). The average score of the all 5 categories of error sources represents each individual's knowledge of error sources.

Table 1: The Iranian TEFL researchers' Error-awareness in measure development process

	Sources of error mentioned in the research narratives	Type	Token	Percentage of participants who are aware of this group of errors	Proportion of observed sources to the expected sources
Individual-related	Idiosyncratic	5	24	76	0.62
	Generic	4	23	45	0.40
Method-related	Item content	3	12	16	0.43
	Response format	4	14	25	0.33
	Administration issue	3	3	10	0.21
Total		19	76		

The patterns of Iranian TEFL researchers' knowledge and sensitivity toward different sources of error in measure development show that generally after constructing the items of the measure they rarely attend to modifying it to reduce the effect of error sources. It seems that construct definition and domain delineation only happens once and at the beginning, and that there is almost no interaction between these two initial phases and the subsequent steps of measure development algorithm. Another important pattern of thought that shows itself in qualitative analysis of the content of the stories is that participants usually see sources of error as compulsory factors that cannot be

controlled, and at best can be reported as limitations of the study. The idea of modifying and correcting the content or format of the developed measures is rarely observed in the narrative episodes allocated to error sources in the research narratives. The analysis of causal links between the subsequent stanzas in the stories most of the time show a forward trend i.e. the researchers tend to go through the steps of measure development one after another without feeling the need for coming back and modifying the status quo to improve the quality of the devised scale. It can be observed that the cyclic nature of the measure development algorithm is not adequately appreciated by Iranian TEFL researchers.

Among the idiosyncratic sources of individual-related error, language difficulties, fatigue, and memory/attention vacillations seem to be more important for researchers. Among generic sources of individual-related error, faking good/bad, acquiescence bias, and rater dispersion bias are more frequently reported. Among item content sources of method-related error, leading questions and direction of wording effects received more attention from the researcher. In the area of response format sources of method-related errors, number of response categories, and leniency/stringency were the most frequent ones. And finally, within administration sources of method-related errors, variations in administration, and distracting setting were more frequently mentioned in the research narratives told by the participants of the study.

5.2. Algorithmic Knowledge and Error Awareness

To calculate the extent to which Iranian TEFL researchers observe the standard algorithm for developing the measures used in their studies, an optimal matching analysis was run. The correlation between algorithmic

knowledge scores and error-awareness scores was analyzed using the Spearman formula. The result of the analysis shows that there is a significant correlation (0.76) between the two sets of scores. The observed relationship is justifiable; a part of the scores for error-awareness is associated with participants' algorithmic knowledge because, as mentioned before, some errors can be reduced through a cyclic approach to the steps in measure development process. Therefore, if someone is more familiar with the subsequent steps in the algorithm, and understands the nature of statistical procedures used in each step, he is more likely to consider the possibility of going back and forth between different stages, and to try to apply his knowledge of error sources to reduce their effect.

5.3. Temporal Markers of Measure Development Knowledge

The results of Cox proportional regression analysis show to what extent the time spent on each of the steps in measure development algorithm can predict participants' knowledge of measure development. The scores of measure development comprise error-awareness scores, and algorithmic knowledge scores. The contribution of each step of the algorithm is shown in Table 2.

Table 2: Proportional model of measure development knowledge in TEFL research: Estimated effects of covariates using a Cox model

	B	SE	Exp (B)
Domain delineation	0.214	0.257	0.327
Item generation	0.316	0.161	0.729
Dimensionality testing	0.456	0.248	**1.578***
Reliability testing	0.398	0.105	1.029
Validity testing	0.493	0.315	**1.637***

The exponents of the B coefficients are reported in the final column of Table 3.

These values show the contribution of the independent variables i.e. steps in error measurement algorithm to the general knowledge of measure development. This makes them more directly interpretable. The events and times that provide the analytic focus of this type of event history modeling introduce some temporal elements that begin to make the analysis more narrative in character. In this context, narrative might be understood as the time spent for each of the steps in the algorithm. According to the results, dimensionality and validity testing can significantly predict one's knowledge of measure development while other steps in the algorithm do contribute to this knowledge in the following order: reliability testing, item generation, and domain delineation.

6. Discussion and Conclusions

Exploration of the Iranian TEFL researchers' knowledge of measure development and error measurement in research design show that error-awareness and algorithmic knowledge of measure development process play a significant role in constituting such knowledge. More importantly, these two types of knowledge are not totally independent of each other. Iranian TEFL researchers seem to be more inclined to look for the sources of error among individual-related factors while ignoring many possible method-related sources of error including item content, response format, and administration issues.

Borg's (2009) study shows that attitudinal, conceptual, procedural, and institutional barriers hamper teachers' engagement in research. The realization of these general categories needs to be diagnosed and explicated in actual research settings. The present study was basically set to accomplish this task.

The qualitative and quantitative (descriptive and inferential) results presented here can be understood in the line of previous research specially the

ones conducted by Borg (2006; 2007; 2008; 2009). Here, we discuss each of Borg's (2009) categories in relation to our finding. According to him, a part of the problems in the TEFL research projects are originated in the researchers' attitude toward concepts such as measuring scale, questionnaire and measurement error. A considerable number of our participants reported that they had not made any attempt to reduce the effect of sources of error in their measurement. Iranian researchers have become accustomed to either ignoring the possible sources of error or merely reporting them as a part of the limitations of their study. This shows that they either see these sources as inevitable realities that accompany any research project or think it is not worth trying to diminish their effect on the result. It seems once they develop their scale and put it through the piloting phase, at best some non-functioning items can be omitted, but the ones which remain in the scale do not need further modifications, in their opinion. This type of attitude toward sources of error in quantitative research seems to have been wrongly compromised. Iranian TEFL researcher' sensitivity regarding measurement error is fading away.

One might argue that this cannot be the case since the number of the researchers who attempt statistical analyses to support the reliability and validity of their results in constantly increasing. However, according to the findings of the present study, the fact is that obsession with statistical indicators after administering the tests and questionnaires cannot replace researchers' sensitivity toward sources of error that can endanger the validity of their results from the beginning, starting the early stages of measure development i.e. domain delineation and item construction. The cyclic nature of measure development, which is a very important issue emphasized by many experts in behavioral measurement and research design (Viswanathan, 2005), is being ignored and practically discarded by many Iranian TEFL researchers. This is

exactly where the researchers' error-awareness and algorithmic knowledge intertwine. Algorithmic knowledge cannot be built up without an adequate knowledge of sources of measurement error. TEFL researchers' attention to the cyclic nature which is expected to be realized in the process of measure development is dependent on their error-awareness. Taking an ignorant approach toward the possible sources of error directs the researchers to a linear path for measure development and administration not a cyclic one.

This attitudinal barrier is affecting a great number of research projects and the validity of their results. Mathematical indicators are being abused to cover for these attitudinal defects. In other words, instead of refining and honing the test items, the researchers administer the tests and then try to validate the results by using a range of statistical procedure one of which would ultimately produce the desired mathematical indicator that can apparently be utilized as a witness to justify the results of the research projects. Our results also show that the attitudinal barrier contributes to another series of problems which are called conceptual barriers (Borg, 2009). When the researcher lose their sensitivity toward the sources of error, constructs assumed by them and consequently the scales developed to measure those constructs may transform in ways hidden to the researcher.

The significant role of dimensionality and validity testing in constituting participants' knowledge of measure development imply that focusing on statistical procedures such as explanatory and confirmatory factor analysis and structural equation modeling (SEM) can help TEFL researchers to expand their research competency to a great extent. The results of our study show that many Iranian researchers avoid using complicated statistical procedures such as SEM (structural equation modeling) only due to procedural and institutional barriers (Borg, 2009). That is to say, because of the high working standards

(demand for a great number of participants, logical complexities, and lack of mathematical knowledge) associated with such statistical procedures, the researcher usually prefer to compromise the validity and usability of the results of their study instead of going though a lot of academic pain. And sometimes the researchers do use high-standard statistical procedures such as regression analysis but are totally ignorant to the mathematical conditions which have to be regarded to ensure the validity of the results. For example, the number of the data items in regression analysis is very important. Given the circumstances of each study, a certain number of participants are required to ensure at least minimal validity for the results. However, not event one participant in the present study reported any though related to the issue of the number of participants. It seems many TEFL researchers feel free to use whatever statistical procedure they wish and are challenged neither by their supervisors nor by the audience community of the academic journals they contribute to.

Given the central importance of measure development in creative and independent research projects in TEFL, researcher education programs need to develop organized syllabi to build up TEFL researchers' knowledge of measure development though expanding their error-awareness and regulating their algorithmic knowledge. While most of the researchers are familiar with the steps involved in error measurement, only a handful are able to incorporate this knowledge into a coherent structure to build up algorithmically oriented research designs. Needs analysis is the first step in such research education programs; this study paves a part of the way in that regard. The quality and effectiveness of quantitative research courses which are held for students of TEFL across Iranian universities can be evaluated based on time and energy they allocate to raise students' error-awareness and to expand their algorithmic knowledge.

References

Abbott, A. (1992). "From causes to events: Notes on narrative positivism", *Sociological Methods and Research*, 20, pp. 428-455.

Abbott, A., and Tsay, A. (2000). "Sequence analysis and optimal matching techniques in sociology", *Sociological Methods and Research*, 29, pp. 3-33.

Allison, D., and Carrey, J. (2007). "What do university language teachers say about language teaching research?", *TESL Canada Journal*, 24, pp. 61-81.

Atay, D. (2008). "Teacher research for professional development", *ELT Journal*, 62, pp. 139-147.

Bagozzi, R. P., and Yi, Y. (1991). "Multitrait-multimethod matrices in consumer research", *Journal of Consumer Research*, 17, pp. 426- 439.

Bardo, J. W., Yeager, S. J., and Klingsporn, M. J. (1982). "Preliminary assessment of format-specific central tendency and leniency error in summated rating scales", *Perceptual and Motor Skills*, 54, pp. 227-234.

Bearger, D., and McAdams, D. P. (1999). "Life story and its relation to psychological well-being", Narrative Inquiry, 9, pp. 69-96.

Borg, S. (2006). "Conditions for teacher research", *English Teaching Forum*, 44, pp. 22-27.

Borg, S. (2007). "Research engagement in English language teaching", *Teaching and Teacher Education*, 23, pp. 731-747.

Borg, S. (2008). "English language teachers' beliefs about research: Perspectives from the Netherlands", *Journal of the Dutch Association of Modern Language Teachers*, 9, pp. 3-13.

Borg, S. (2009). "English language teachers' conceptions of research", *Applied Linguistics*, 30, pp. 358-388.

Brown, J. D., Knowles, M., Murray, D., Neu, J., and Violand-Sanchez, E. (1992). *The Place of Research Within the TESOL Organization,* Cambridge: TESOL.

Connelly, F. M., and Clandinin, D. J. (2006). "Narrative inquiry", in Green, J. L. n., Gamilili, G., and Elmore, P. B. (eds.), *Complementary methods for research in education*, NJ: Erlbaum, pp. 375-385.

Craig, C. J. (2007). "Story constellations: A narrative approach to contextualizing teachers' knowledge of school reform", *Teaching and Teacher Education,* 23(2), pp. 173-188.

Crites, S. (1971). "The narrative quality of experience", *Journal of American Academy of Religion,* 399, pp. 291-311.

Czarniawska, B. (2004). *Narratives in Social Science Research,* London: Sage.

Elliot, J. (2005). *Using Narrative in Social Research: Qualitative and Quantitative Approaches,* London: Sage Publication.

Franzosi, R. (1998). "Narrative as data: Linguistic and statistical tools for the quantitative study of historical events", *International Review of Social History,* 43, pp. 81-104.

Gee, J. P. (1986). "Units in the production of narrative discourse", *Discourse Processes,* 9, pp. 391- 422.

Ghiselli, E. E., Campbell, J. P., and Zedeck, S. (1981). *Measurement Theory in the Behavioral Sciences,* Sac Francisco: Freeman.

Lang, R. R. (2001). "Story grammars: Return of a theory", in Mateas, M. and Sengers, P. (eds.), *Narrative intelligence: Advances in consciousness research,* Philadelphia: John Benjamins, pp. 199-212.

McDonough, J., and McDonough, S. (1990). "What's the use of research?", *ELT Journal,* 44 (2), pp. 102-109.

Nunan, D. (1997). "Developing standards for teacher-research TESOL", *TESOL Quarterly,* 31, pp. 365-367.

Ochberg, R. L. (1996). "Interpreting life stories", in Josselson, R. (ed.), *Ethics and process in the narrative study of lives,* CA: Sage, pp. 35-49.

Viswanathan, M. (2005). *Measurement Error and Research Design,* London: Sage.

Xu, Y., and Liu, Y. (2009). "Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience", *TESOL Quarterly,* 43, pp. 493-513.