Substantive Scale Development: How to Design, Administer, and Verify a Likert Scale Questionnaire for a Research Project

David Ockert

Tezuka Corporation

Abstract

Language researchers face the task of creating an instrument capable of actually gathering the information that they want to analyze and report on. This paper explains how to conduct research by reporting the author's development, piloting, administration, and analysis of a substantive scale survey for research purposes. A substantive scale uses questions and a scale system (e.g. a Likert scale) to gather data for analysis. This paper follows the outline of a research paper and the sections are explained using the author's own research project to measure student (N=104) intrinsic and/or extrinsic motivation. Exploratory factor analysis confirmed the students tended to adhere to one or the other motivational type. The author hopes that readers will gain an understanding of the research process, including key terms and definitions, and proceed with greater confidence to design their own research projects and share their results with the greater community interested in language learning and research.

Introduction

To begin a research project survey, important questions must be answered: How to formulate a hypothesis? How to write questions and

how many? And in what order? What type of scale should be used? For example, the most common type of scale is the Likert scale, which uses a 1 to 5, 6, or 7 numeric system, which corresponds to a series of answers such as always, sometimes, never, etc. There are more than a dozen different scale types (Alreck & Settle, 2003); however, sometimes a simple Yes/No question will suffice (Stone, 2003).

What are good sources of information to read before designing the survey? Not only are the statistical methods quite complicated and intimidating, but the terminology itself can be difficult to comprehend. Therefore, before beginning, have a clear goal. For this research project, I wanted to research (1) Japanese student motivation, (2) their pedagogical activity preferences, and (3) the relationships between the two. It would have been possible to administer a survey by another author, known as replication, but after a brief literature review, I discovered that making one's own survey to suit the learning environment was preferable (Dörnyei, 2001).

Next, I decided on the data analysis options before the development of the survey, since the survey data will determine what type of analysis is appropriate. (Brown, 2001; Dörnyei, 2001). So, determine the goal and decide how to get there, and then write the survey questions.

Finally, what to do with research findings? Most teacher/researchers hope to publish their findings in a professional journal. In fact, having a proven record of publications on a job applicant's curriculum vitae can make the difference in landing a job or not (McCasland & Poole, 2004; McCrostie, 2007). There are many journals to choose from within JALT. We have OnCUE Journal, The JALTCALL Journal, the JALT Hokkaido Journal, The Language Teacher, the JALT Journal, and more. A careful reading of their respective submissions guidelines will help get the research write-up started on the right foot.

Terminology and Definitions

The terms substantive scale, instrument, and survey are synonymous. An item refers to a question on the survey. The items attempt to measure a construct, which actually refers to a way of thinking that exists within the minds of the participants (Brown, 2001). Since we cannot actually see a construct, we must test for it by attempting to measure it. Therefore, items are created to measure the construct.

Once the survey has been administered and the data collated, the accumulated result of the individual test item is referred to as a variable. A variable is the quantified means to measure an observable characteristic of a phenomenon (Voelker et al, 2001). For example, item number five on a questionnaire administered to one hundred students will have a total of one hundred responses measuring the target construct. These responses are all compiled and analyzed using any number of statistical methods and the result is the variable for item five, such as the average, derived from the responses from all of the participants. From then on, the research may simply refer to variable five for the now numerically operationalized data from survey question five (for more information, please see Brown, 2001, pp. 16-17).

Before Scale Construction: Background Reading

First, decide on an area in our field that is interesting. Next, begin reading the relevant literature. The literature review serves three purposes (1) to see if the research question(s) have already been answered, (2) to provide the reader with the necessary background concepts, and (3) how this prior research supports the present research endeavor. How has the work done by other researchers lead up to the research questions? The survey used as an example for this article was designed to test student motivation; therefore, I read publications by Dörnyei (2001), Tremblay and Gardner (1995), Gardner and Lambert (1959), and others. Since the learners in this research were primarily

from English as a second language (ESL) learning environments, and this article's questionnaire was designed for Japanese students in an English as a foreign language (EFL) environment, this was taken into consideration in constructing the instrument (Dörnyei, 2001).

Research Questions and Hypotheses

Much research on motivation of ESL learners has been reported on, particularly for French Canadian learners (Gardner & Lambert, 1959; Tremblay & Gardner, 1995; Noels, et al., 2000). Regarding extrinsic, intrinsic, and amotivation (unmotivated) motivational orientations of these learners, experts generally accept that these constructs are not categorically different, but rather exist along a continuum of self-determination (Deci & Ryan, 2002; Noels, et al., 2003). Research by Ryan and Connell (1989, cited in Deci & Ryan, 2000, p. 73) tested for "different types of motivation, with their distinct properties" to confirm that they do indeed "lie along a continuum of relative autonomy." Since this notion could help us learn more about our learners as people, this theory has tremendous value for educators. Intrigued, I wanted to know if the motivation of students in my classes could be divided between an intrinsic and an extrinsic motivational orientation, and wrote items that are hoped to be one or the other (see Appendix; Ockert, 2005; 2007).

Questionnaire Construction

Questionnaire items

Brainstorming works well to get started on item writing (Griffee, 1999). Also, by reading the instruments constructed by the authors mentioned previously, how statements are worded became clearer. Constructing an instrument to measure a group of learners' motivational attitudes toward language learning remains difficult; therefore, when choosing questionnaire statements for a survey there are some rules to

keep in mind according to Stone (2003). These include:

1. Avoid factual statements.

2. Do not mix past and present. Present is preferred.

3. Avoid ambiguity.

4. Do not ask questions that everyone will endorse.

5. Keep wording clear and simple.

6. Keep statements short and similar in length.

7. Express only one concept in each item.

8. Avoid compound sentences.

9. Assure that reading difficulty is appropriate.

10. Do not use double negatives.

11. Do not use "and," "or," or lists of instances. (p. 288)

Furthermore, in Teaching and researching motivation, Dörnyei lists several items from his research, providing a wealth of ideas. Following the advice above, I began writing the survey items following all of the above guidelines.

Next, how many questions are enough to test the hypothesis? This depends on the type of data analysis used. Factor analysis, a complicated statistical procedure, provides groupings of similar question responses to test for possible relationships between specific variables. Finding relationships between variables with factor analysis helps create a stronger and more valid instrument after removing items that do not "fit in." First, start with more questions than may be necessary and discard those that don't correlate well with each other. Working with my MEd professors, we created the survey statements using the expert rating approach (Brown, 2001, pp. 179-80). The first eight are hoped to measure intrinsic motivation and the latter eight extrinsic motivations.

Survey Organization and Item Selection

How much information should be written on the questionnaire? It's

best to keep the instructions clear, simple, and concise. The example survey has two parts: the first eight items test for an intrinsic motivational orientation and the second eight test for an extrinsic orientation. If the respondents perceive a difference in the two sections and indicate answers differently than if the items were arranged randomly, this would result in a response bias. This occurs when the answers given do not reflect the students' true beliefs as a result of the wording or ordering of questions. Therefore, care should be taken to avoid presenting the questions in a manner that has a "pattern" in order to avoid collecting biased data (Gendall & Hoek, 1990). For this reason, the sixteen questions should be arranged randomly on the actual survey.

Rating Scale Formats and Representative Samples

When using a Likert scale, consider what kind of Likert scale will work best. Originally designed by Rensis Likert (1932), this scale usually consists of four, five, six, or seven points. However, there are advantages and disadvantages to not only the number of choices, but also whether or not the number of points is odd or even. For example, the advantage of an evenly numbered scale is that it removes the neutral answer option, which would tell us nothing regarding a positive or negative attitude toward the survey question or statement (Stone, 2003).

It is necessary to view the item from the perspective of a respondent, and the responses from the pilot testing stage need to be examined carefully. If the item is easily answered with a dichotomous option, then a simple Yes/No option should be available instead. For example, if the answers converge on 1 (never) or 5 (always), then the other options of the middle 2, 3, and 4 need not be made available since they would yield little analytical value. It's best not to construct a large number of questions and assume from the start that every item will fit into a standard five or seven point scale (Stone, 2003). This requires careful analysis to understand the underlying item data that compose

the variable data.

Furthermore, Stone says rating scales should follow a graded response format such as never, sometimes, frequently, always; or none, some, a lot, and all. While this may seem easy, the terms used can actually be ambiguous: what is the difference in meaning between usually and frequently? Do the terms none and always mean absolutely and without exception? The meaning will differ according to how each participant uses the terms in everyday life. Therefore, care must be given to make rating scales can be made that solicit information without confusion (Stone, 2003). Finally, depending on the analytical method used to sift through the data, what is a minimum number of respondents necessary to have a representative sample? Most experts agree that twenty randomly selected surveys per 1,000 potential respondents are acceptable. Brown suggests 28-30 as being sufficient (Brown, 2001).

Methods

Pilot testing

The survey should be pilot tested with a small sample group before using it for research purposes. Asking native speakers (NS) to review the questions first will assure that the instrument items make sense to your peers; therefore, ask colleagues to review the questionnaire items beforehand. Any ambiguities in the instructions should be found during the pilot phase and corrected. Researchers may also wonder: What about translating the survey into the respondents' mother tongue (L1)? Or should the items be written in both the second language (L2) and the L1? (e.g., English and Japanese). How can researchers handle issues of low L2 proficiency? Certainly it is a good idea to ask a small representative sample group of non-native speakers (NNS) to check the instrument for clarity. Any problem areas that are difficult

to comprehend should be corrected and re-checked (Griffee, 1999) as was done for this example survey.

Students

The students (*N* = 104) who took this survey were members of my Communication I class in a private university in Japan. This means they are a sample of convenience and the results may not be applicable to the general population of Japanese university students (Brown, 2006). However, approximately 95% of the respondents were male, and the results may be generalized to their peers, although gender was not considered when analyzing the results. Participation was voluntary, anonymous and had no influence on student grades. Therefore, teacher bias and external validity (see below) were eliminated as negative influences since the respondents were all my students in the same environment.

Statistical Analysis

The Statistics Package for the Social Sciences, version 13 (SPSS13) was used for data analysis. Calculating the average (the mean), determining the most frequent response (the mode), and determining the central cut-off point (the median) are commonly used processing methods (Brown, 2001). These are the simplest methods of reporting data. Depending on what information the researcher wishes to report, there are more sophisticated procedures such as factor analysis (see below). When providing the information on the number of respondents, use N for number of respondents; use SD for standard deviation; and for mean, mode, or median, clearly indicate with an M for just one of the three, or write out the word otherwise (Kachigan, 1991).

Factor Analysis

In order to find underlying relationships between the variables, a

multivariate statistical calculation known as factor analysis can be used. Exploratory factor analysis can answer the question, "Will the variables fit together as hypothesized?" (Nunan, 1992; Brown, 2001). Factor analysis will organize the responses in variable groups and analysis of these groups will yield the answer. The survey for this project was created with the hypothesis that the first eight items measure intrinsic motivation and the second eight measure extrinsic motivation. Ideally, they should "cluster together" in two sets of eight. These "clusters" are referred to as "factors" and the author gets to name them. The responses clustered together nicely as hoped (see Ockert, 2005; 2007). For a simple explanation of factor analysis and how it works please visit: <http://www.janda.org/workshop/factor%20analysis/factorindex. htm>

Reliability

As important as the validity of the instrument is the reliability. Does the instrument measure what it purports to measure in a consistent manner at different times? (Brown, 1988; 2001; Griffee, 1999). In other words, do different groups of persons who answer the survey give similar responses? To test the reliability of the instrument the researcher uses the split-half method known as Cronbach's alpha coefficient of reliability (for more information, see Brown, 2001).

Validity

According to Brown (1998; 1996; 2001) and Nunan (1992), there are several types of validity and ways to test them. We will look at the three most commonly referred to types here: internal validity refers to whether or not the questionnaire is in fact measuring what it claims to measure; external validity refers to whether or not those persons taking the survey by answering the questionnaire did so under similar conditions; finally, Brown (2001) explains construct validity as

the "degree to which the survey can be shown experimentally to be measuring whatever construct you are trying to measure" (p.181).

Conclusions

As Griffee (1999) has noted, validation of a survey instrument requires months if not years before administering it for research results. It is a specialized business and should not be taken lightly (Nunan, 1992). However, the dedicated pursuit of proof to a hypothesis remains a worthy goal and provides the foundation for growth and learning in our field, and statistical analysis can help even those of us who are novices gain a better understanding of language learners (Ockert, 2008). Since getting published remains a vital need for most educators, there is no better time to start than the present. Good luck!

Acknowledgement

<p style="text-align:center">*   *   *</p>

David Ockert was born in Michigan, USA. His research interests range from student motivational orientation, learning strategies, and their relationships between specific classroom activities, either traditional or task-based, to educational system development. He has a dual-major BA in Political Theory and Constitutional Democracy (PTCD) and East Asian Studies (Japanese) from James Madison College, Michigan State University, and a MEd from Temple University Japan. He can be contacted at davidockert1@gmail.com

References

Alreck, P. L., & Settle, R. B. (2003). T*he survey research handbook*. New York: Mcgraw Hill.

Brown, J. D. (2006). Generalizability from second language research samples. *Shiken, 10*(2), 24-27.

Brown, J. D. (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.

Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, New Jersey: Prentice Hall Regents.

Brown, J. D. (1988). *Understanding research in second language learning*. Cambridge: Cambridge University Press.

Deci, E. L., & Ryan, R. M. (2002). Overview of self-determination theory: An organismic dialectical perspective. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research*. New York: University of Rochester Press.

Dörnyei, Z. (2001). *Teaching and researching motivation*. London: Pearson, ESL.

Gardner, R. C., & Lambert, W. E. (1959). Motivational variables in second language acquisition. *Canadian Journal of Psychology, 13*, 266-272.

Gendall, P., & Hoek, J. (1990). A question of wording. *Marketing Bulletin*, (1), 25-36. Retrieved February 3, 2009 from http://marketing-bulletin.massey.ac.nz/V1/MB_V1_A5_Gendall.pdf

Griffee, D. T. (1999). Questionnaire construction and classroom research. *The Language Teacher, 23*(1). Retrieved December 20, 2008 from http://www.jalt-publications.org/tlt/articles/1999/01/griffee

Kachigan, S. K. (1991). *Multivariate statistical analysis: A conceptual introduction*. New York: Radius Press.

Likert, R. (1932). A technique for the measurement of attitude. *Archives of Psychology, 140*, 1-55.

McCasland, P., & Poole, B. (2004). The person, the package, the presentation: Lessons from a recent job hunt. T*he Language Teacher, 28*(10). Retrieved December 20, 2008 from www.jalt-publications.org/tlt/articles/2004/10/mccasland

McCrostie, J. (2007). Perishing to publish: An analysis of the academic publishing process. *OnCue Journal, 1*(1), 74-79. Retrieved February 3, 2009 from jaltcue-sig.org/node/57

Noels K. A., Pelletier L. G., Clément R., & Vallerand, R. J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning, 50*, 57-85.

Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.

Ockert, D. (2008). Statistical analysis of an online survey questionnaire of student motivation and classroom activity preferences. Paper presented at the June 22 meeting of the Shinshu Chapter of the Japan Association for Language Teaching (JALT), Matsumoto, Japan.

Ockert, D. (2007, Spring). Computer assisted learning about language learners' learning styles. *IATEFL CALL Review*, 17-20.

Ockert, D. (2005). Student motivation and pedagogical activity preferences. Retrieved March 16, 2006 from http://jalt-publications.org/archive/proceedings/2005/E151.pdf

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. New York: Lawrence Erlbaum.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Stone, M. (2003). Substantative scale construction. *Journal of Applied Measurement, 4*(3), 282-297.

Tremblay, P. F., & Gardner, R. C. (1995). Expanding the motivation construct in language learning. *Modern Language Journal, 79*, 505-520.

Voelker, D., Orton, P., & Adams S. (2001). *CliffsQuickReview Statistics*. New York: Wiley.

Appendix

Sample Survey

What is your attitude toward learning English? Circle the number of the answer that best matches your opinion:

1 = strongly disagree   2 = disagree   3 = neutral   4 = agree   5 = strongly agree

1) I enjoy studying English.	1	2	3	4	5
2) English is important to me because I want to make friends with foreigners.	1	2	3	4	5
3) English is important to me because I want to study overseas.	1	2	3	4	5
4) English is important to me because I want to read books in English.	1	2	3	4	5
5) Language learning often makes me happy.	1	2	3	4	5
6) Language learning often gives me a feeling of success.	1	2	3	4	5
7) I study English because being able to use English is important to me.	1	2	3	4	5
8) English is important to me because I like English movies or songs.	1	2	3	4	5
9) I study English because it will make my teacher proud of me/ praise me.	1	2	3	4	5
10) I study English because it will make my parents proud of me/ praise me.	1	2	3	4	5
11) I study English because I want to do well on the TOEIC test.	1	2	3	4	5
12) I study English because I want to do well on the TOEFL test.	1	2	3	4	5
13) In the future, English will be helpful/ useful to me.	1	2	3	4	5
14) English is important to me because I might need it later for my job.	1	2	3	4	5
15) I study English because all educated people can use English.	1	2	3	4	5
16) I study English because I must study English.	1	2	3	4	5