

# The learner capital questionnaire: Some preliminary observations

Jean-Pierre Joseph Richard

## Introduction

Richard (2013) outlined the initial development of a tool to measure learner capital, a composite of cultural and learning capital, as well as life experiences and goals. Cultural capital are non-financial assets which contribute to social mobility, such as linguistic abilities, physical objects, and educational credentials (Bourdieu, 1986). Learning capital refers to competencies—skills (i.e., learning habits) and attitudes towards learning (e.g., eagerness to learn) (Kariya, 2010). Richard argued that learners bring with them *deposits* of both  $\pm$ cultural capital and  $\pm$ learning capital into their academic environments, as well as  $\pm$ nested life experiences (e.g., overseas study) and  $\pm$ distal goal orientations; the sum total of which he labelled learner capital.

A problem that occurs in our field is that instruments whether developed to measure a certain linguistic skill or individual difference or other often lack validity (Apple, 2011; Elwood, 2011). Thus, this paper attempts to provide further rationale for the learner capital questionnaire. Another goal of the paper is attempt to answer the following question: How reliable are responses? A final goal of this paper is identify preliminary results from certain sections of the questionnaire.

A number of general research specialists (including Coste, Guillemain, Pouchot, Fermanian, 1997; Cronbach, 1990; Field, 2009; Netemeyer, Bearden & Sharma, 2003; Smith, 2001; Trochim & Donnelly, 2008) and academics in

our field (including Apple, 2011; Beglar, 2010; Elwood, 2011; Gardner & MacIntyre, 1993; Sakui & Gaies, 1999) have discussed general characteristics of psychometrics. Measurement tools need to be standardized and norms need to be established (Netemeyer, Bearden & Sharma, 2003); and this can occur when measurement rules are explicitly described, the practicality of the tool is considered, and results depend on the participants, not the measure's administrators (Nunnally & Bernstein, 1994). In the section below, I discuss measurement validity and reliability—two fundamental characteristics of psychometrics.

In research, construct validity, separate from reliability, refers to the extent to which a concept corresponds accurately to the real world. Validity, in measurement, then is the degree to which a tool measures what it purports to measure; that is, the degree to which researchers can make well-founded inferences from the operationalizations in their study to the theoretical constructs on which those operationalizations are based (Trochim & Donnelly, 2008). As an example, a test which purports to measure a construct labelled general listening skills, but which focuses primarily on testing aural comprehension of individual lexical items is not measuring what it purports to measure. Construct validity is then how well one researcher's questionnaire (the lower portion of Figure 1) matches her theory (the upper portion of Figure 1).

Trochim and Donnelly differentiated between two types of validity: translation validity, which refers to how well you translated the idea of your measure into its operationalization; and criterion-related validity, which refers to how well the measure correlates with other independent constructs. See Table 1.

Face validity refers to the translation of the construct. On its face, does the questionnaire look like it is measuring what it says it is measuring. Although this requires subjective judgement, Trochim and Donnelly suggest that one way to improve face validity is to have several experts review the measure to

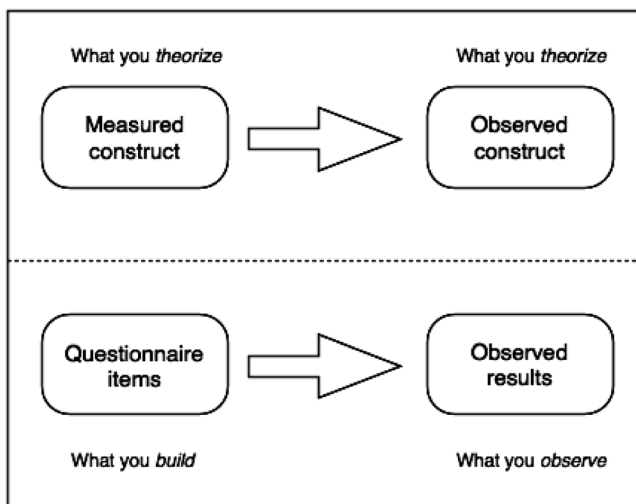


Figure 1. Construct validity (adapted from Trochim and Donnelley (2008).

offer their judgement. As an example, does a test which claims to be a new L2 test of vocabulary breadth appear to measure vocabulary breadth? Content validity requires that a measure is checked against relevant content domain. In other words, are there similar items on other measures and does each item measure what it is intended to measure?

With predictive validity, the measure is assessed by how well it predicts that which it theoretically should predict. Thus, using the same example as above, a measure of vocabulary breadth should be able to predict in part reading ability as breadth of vocabulary is an important component of L2 reading. Concurrent validity assesses how well the measure discriminates between two different groups of people. Thus, a new vocabulary breadth test should separate people by their ability; those with high scores should indeed have higher ability, and those with lower scores should have lower ability. With convergent validity, researchers are interested in how similar their new test is with other, theoretically similar measures. Thus, the scores from a new

vocabulary size test should correlate with the scores of similar vocabulary size measures. Discriminant validity assesses how dissimilar one measure is to other theoretically dissimilar measures. Thus, scores from the new vocabulary size test should not correlate with scores from an L1 biology test, because the two constructs are theoretically dissimilar. Lastly, Trochim and Donnelly (2008) caution that both convergent and discriminant validity need to be both working to establish construct validity.

Table 1. Construct validity according to Trochim and Donnelly (2008).

Translation validity	Criterion-related validity
Face validity	Predictive validity
Content validity	Concurrent validity
	Convergent validity
	Discriminant validity

In the human sciences, some researchers take a definitionalist approach when defining constructs. That is, a construct they wish to operationalize is so accurately defined, that its operationalization is simple. As an example, a vocabulary size test is or is not measuring what it purports to measure. However, most researchers take a relationalist approach. That is, constructs are rarely measured in black and white terms. A tool, thus for relationalists, is measuring, more or less, what it is intended to measure. As an example, a test that claims to measure L2 reading ability may likely be also measuring, in part, other constructs, such as vocabulary breadth and memory because of partial overlap of the constructs. In the simplified Figure 2, overlap between different constructs and L2 reading test can be seen. In Figure 3, a similar, simplified figure is used to show learner capital overlap.

Reliability is the degree to which a test consistently or dependably returns the same score (Gay, 1987; Nunnally, 1978; Trochim & Donnelly, 2008). A general model of measurement reliability is:

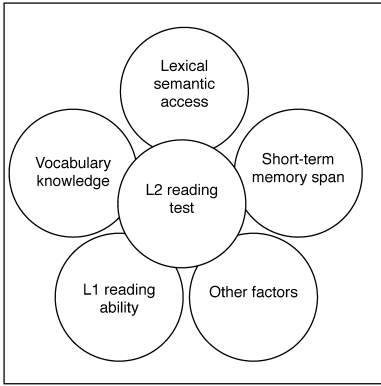


Figure 2. Simplified representation of a more or less relationalist definition of an L2 reading test.

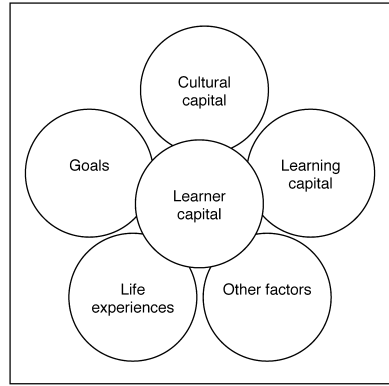


Figure 3. Simplified representation of a more or less relationalist definition of learner capital.

$$X = T + e$$

That is, the observed score ( $X$ ) is composed of two components, which are true ability ( $T$ ) plus random measurement errors ( $e$ ). What researchers are able to observe is on the left side of the equation, the observed score. Error is sometimes divided into two components (Trochim & Donnelly, 2008): random error (e.g., mood of the test takers), and systematic error (e.g., traffic conditions near the test-taking site), resulting in the following equation:

$$X = T + e_r + e_s$$

Trochim and Donnelly (2008) suggest several ways to reducing measurement error. These include pilot-testing the measurement, double-checking data on both paper and computer files, and data triangulation. There are a number of ways to calculate reliability, such as test-retest reliability, parallel forms, and internal consistency.

As an example of reliability, a test which purports to measure vocabulary breadth is taken by one student once a day over five days. The student scores each time five wildly different scores such as 1500, 4800, 750, 3600, and 2900.

This test would not be reliable, because the scores are not consistent or dependable. Reliability is a required feature of psychometrics, but it is not sufficient in itself. For example, another student may score 1800, 1800, 1800, 1800 and 1800 on a measure which claims to be measuring vocabulary breadth. The test appears to be reliable because the student's score is consistent and dependable. However, if this student has a true vocabulary breadth score of 2800, the test cannot be said to be valid. Thus, validity is also an important and necessary feature of psychometrics.

### **Hypothesis**

The hypothesis states that the two groups of students that participated in this small research can be separated, as expected *a priori*, based on a small number of learner capital items discussed below. Doing so will help to provide, in a small way, validation of the learner capital questionnaire.

### **Methodology**

#### **Participants**

The participants described in this paper come from a larger, year-long doctoral dissertation study. The total sample of participants in the larger dissertation research project is over 1000. However for this small paper, two complete classes were selected for comparison. The first group of participants, all Japanese, ( $n=25$ ) are from a national university (NU) with a *t*-score (hensachi) of 74. This group includes one female and 24 males. Of these 25 students, 23 are first-year students, and two are second-year. The participants in the larger dissertation research project also completed the Vocabulary Size Test (Nation & Beglar, 2007). The average score of the NU students on the Vocabulary Size Test was 31.92, equivalent to approximately 3200 known words in English. The second group of participants, all Japanese, ( $n=25$ ) are from a private, female junior college (JC) with a *t*-score of 45. Of these 25 students, 24 are first-year students, and one is second-year. Their average Vocabulary

Size Test score was 20.40, equivalent to approximately 2040 known words in English. The difference in vocabulary size scores is significant with a very large effect size:  $t(48)=10.94$ ,  $p=<0.0001$ ,  $d=3.16$ . Thus at the onset, the two groups of participants are different in several important ways: one group attends a top-ranked university, has significantly higher vocabulary size scores, and is predominantly male; while the second group attends a below-average-ranked junior college, know significantly fewer English vocabulary items, and is female-only.

### Materials

Although all participants completed all sections of the learner capital questionnaire (Richard, 2013), only results from questions 1 through 15 are reported here. These questions are:

1. What is your gender?
2. What year are you at university?
3. What was your high school *t*-score?
4. How many years were you *ronin* (students who have finished high school but have yet to enter university) before entering this university?
5. Do you have a part-time job at the moment?
6. What previous experiences have you had with English?
7. What previous experiences have you had studying for other academic certificates or participating in cultural and sporting contests?
8. There were four questions related to L1-reading. These were scored on a 6-point Likert-scale ranging from 1 (completely disagree) to 6 (completely agree). They were:
  - a. I read only when I need to.
  - b. Reading is a favorite hobby of mine.
  - c. I read books from start to finish.

d. Reading is a waste of time.

There were six questions, also scored on the same 6-point Likert-scale as above, related to university and future. These were:

9. This is my first-choice university.
10. This is the university my parents wanted me to attend.
11. My parents want me to have the same type of employment as they have.
12. My parents want me to have the same level of education as they have.
13. My parents want me to have a higher level of education as they have.
14. My parents were eager for me to study when I was growing up.
15. Finally, there were two questions related to future goals.
  - a. What is the final level of education you will attain?
  - b. What career will you have? Be specific.

### **Procedure**

The questionnaires were distributed by colleagues in their English-language general English courses. The entire learner capital questionnaire takes approximately 20 minutes to finish. These 15 questions take approximately five minutes to complete. For the most part, participants respond to each item by circling an answer from a list of choices given. A few questions, however, require written responses. For these, participants wrote in Japanese. All responses, marksheet-type answers and Japanese written responses, were input into a spreadsheet document by two trained research assistants.

### **Analyses**

As this is only a preliminary introduction to a small section of the learner capital questionnaire with a very small number of participants, raw score totals, percentages, average scores for items, and *t*-tests were calculated. These were calculated using standard functions within a spreadsheet document.



Correlation coefficients were calculated. To do so, dummy codes were assigned for present university (1=JC; 2=NU) and final educational goal (1=junior college; 2=university; 3=graduate school). A multiple-linear regression model was built to investigate how the learner capital questionnaire could predict the vocabulary size test results. Finally, a reliability estimate (Cronbach's  $\alpha$ ) was calculated.

## **Results and Discussion**

As described above, participants came from two different samples: one group attend a high-level national university (NU); the other attend a low-ranked junior college (JC). Thus, on many of the items below, it is expected that these populations will differ. This paragraph describes results of the *t*-score of the participants high schools, the number of years the students were ronin before entering university, and their current part-time work experience. First, the average high school *t*-score (Q3) of the students who attend NU was 58 (although, unknown for four students) compared with 51 (unknown for one) for JC students. Secondly, considering the number of years students are ronin (Q4), 36% of NU students were ronin for at least one year compared to only 4% of JC students. This later figure is slightly smaller than the national percentage (5%) of high school graduates who become ronin each year (Sugimoto, 2010); whereas the former figure is seven times larger than the national average indicating the competitive nature of the NU entrance examination. Lastly, regarding part-time work (Q5), 24% of the NU participants were working part-time when the questionnaire was completed, and they averaged 12 hours per week of part-time work; compared with 64% of the JC students who averaged 20 hours per week. On these three questions, the groups of students appear to be different, the group of students from NU went to higher-ranked high schools, were much more frequently ronin, and worked less per week compared with the JC students.

Next, there were five questions (Q6a, Q6b, Q6c, Q6d and Q6e) related to experiences using English. Of the NU students, one participant had previously attended an English conversation school, two had attended an English-language cram school, one had experienced an overseas homestay, and one had entered an English speech contest. Of the JC students, five participants had attended an English conversation school, nine had attended an English-language cram school, one had experienced an overseas homestay, and two had studied in an overseas language school. On these questions, there were few examples of experience with English; although, overall it appears the JC students have had more experiences with extra-curricular English study. In particular, the number of students from JC (36%) who have attended a cram school to learn English is strikingly large compared with NU (8%). The vocabulary size scores from the JC students indicate a smaller vocabulary compared with the NU students; and this smaller vocabulary size may be an indication of poorer English language skills compared with the NU students; and thus, the JC students may have been attending a cram school simply to catch up with their peers.

The next set of questions ask about other academic certificates (Q7a) that students had studied for and various activities and contests that students had participated in (Q7b). From the NU students, 68% had received academic certificates, and of these the largest number ( $n=14$ ) were for Japanese language tests (漢字検定), and the others were math (数学検定) ( $n=3$ ) and abacus (そろばん検定) ( $n=4$ ). Also, from this group, four participants indicated they had two or more certificates. From the JC students, 88% had received academic certificates, of which the largest number were for Japanese language tests ( $n=16$ ). A large number of students also selected other ( $n=7$ ), which included service-oriented certificates such as aromatherapy and massage. Two students indicated they had math certificates, and one each indicated calligra-

phy and accounting. Finally, from this group, seven participants indicated they had two or more certificates. Regarding competitions, 36% of the NU students had previously entered competitions. These were: Japanese speech contests ( $n=2$ ), sporting event ( $n=3$ ), art or music-related events ( $n=3$ ) and other ( $n=2$ ). This compares with 48% of the JC students, of which five entered sporting events, three each entered Japanese speech contests and art or music-related events, and other. From these results, although the numbers are small, it seems that students from JC had more opportunities to partake in extra-curricular events, particularly cultural related. This may be unsurprising as Yamamoto and Brinton (2010) noted that Japanese women have more cultural capital than Japanese men, and that in Japan, cultural capital has a less important effect on academic achievement than parental socio-economic status.

The next four questions (Q8a, Q8b, Q8c, and Q8d) were Likert-scale questions about L1-reading on a 6-point scale from completely disagree to completely agree. For the first question, “I only read when I need to”, the average score of the NU students was 3.40 compared with 4.28 for the JC students. A test of statistical difference ( $t$ -test) and effect size (Cohen’s  $d$ ) were used:  $t(48) = 2.37$ ,  $p=0.02$ ,  $d=.68$ . This indicates that the difference between the populations were statistically different and the size of the effect was moderate to large (Cohen, 1988). For the remaining questions below in section 8, then 9 to 14, results from  $t$ -tests and Cohen’s  $d$  are shown when the group differences were found to be statistically different. For the second question, reading is one of my hobbies, the average score for the NU group was 3.32 compared with 3.16 for the JC group. For the third question, when I read, I always need to finish what I started, the NU group scored 4.16 compared with 4.36 for the JC group. For the last question, reading is a waste of time, both groups scored well below the mid-point. The average score for the NU group was 1.48 com-

pared with 1.96 for the JC group. Of the four questions in this section, results from one showed statistical differences between the two populations studied, and the effect size was moderate to large.

The following six questions were related to university choice, education level and future career. The first of these questions asked the participants if they were attending their first-choice university (Q9). For the NU group, the average score was 5.32 compared with 2.96 for the JC group.  $t(48)=5.10$ ,  $p<.0001$ ,  $d=1.44$  (indicating a very large effect size). The next question asked if the university the students attend was the university their parents wanted them to attend (Q10). For the NU group, the average score was 4.56 compared with 3.60 for the JC group. This difference was statistically significant:  $t(48)=2.38$ ,  $p=0.022$ ,  $d=.69$ . The third question asked if the parents hope the students have the same type of employment in the future (Q11). For the NU group, the average score was 2.08 compared with 2.12 for the JC group. The fourth (Q12) and fifth (Q13) questions asked if the participants parents want them to have the same or higher level of education. For the NU group, the mean score for these questions were 2.64 (same) and 4.40 (higher) compared with 2.68 (same) and 4.12 (higher). The last question (Q14) asked if the participants parents had been eager about their education as they were growing up. For the NU group, the mean score for these questions were 4.04 compared with 3.79. In this section, of the six questions, two questions revealed statistical differences between the two populations studied and the effect size was moderate-to-large for one question and very large for the other.

The final two questions reported in this small article asked about the participants future education (Q15a) and career goals (Q15b). Regarding the first question, 84% of the respondents from NU indicated they intended to graduate from their current university and 16% were going to complete graduate work. This compares with 64% of the JC students who indicated they

would graduate from their current junior college, while 36% indicated they hoped to transfer to a university to complete their studies. Regarding career goals, while only 24% of the NU students responded, their careers would be: national civil servant, United Nations employee, consultant, management-level staff, a CPA and an economic policy researcher. From JC, 64% of the participants responded. Their intended future careers are: hotel and airport staff ( $n=7$ ), company employees ( $n=7$ ), and two each for office clerks, English-language related and junior high school teacher. From this, it is clear that the intended final levels of education are different, with NU students completing more years at higher-ranked universities; and the participants subsequent future career goals reflect this with more academic and managerial positions for the NU students and more service-oriented positions for the JC students.

Translation validity refers to how well the construct has been operationalized. Richard (2013) reported that Japanese students, doctoral dissertation colleagues and academics reviewed the items on the learner capital questionnaire, and they indicated that on its face, many of the items of the learner capital questionnaire appear to relate to cultural capital, learning capital, life experiences and goals, which are the sum components of my definition of learner capital. Likewise, as similar items can be found in related questionnaires, the items discussed in this paper and in Richard (2013) indicate there appears to be content validity as they are measuring something labelled *more or less* learner capital. Thus, tentatively, I claim that the learner capital questionnaire has translation validity.

How well the measure correlates with other independent constructs is criterion-related validity. For example, students high in learner capital should be at higher, more prestigious universities; have parents with more education, more prestigious job, have stronger goal orientations. Table 2 displays

Table 2. Table of Correlation Coefficients.

UniT-s	VST	HST-s	Ronin	8a	8b	8d	9	10	11	12	14	15
	0.85**	.57**	.39**	−0.32*			0.59**	0.32*				0.71**
		.50**		−0.31*		−0.33*	0.52**					0.49**
							.45**				.31*	.38**
							−.281*					.36*
												−.30*
						−.55**	.46**					
							−.56**					
									0.49**			
										0.37**		

\*\* Correlation is significant at the 0.01 level (2-tailed). \* Correlation is significant at the 0.05 level (2-tailed).

Note. UniT-s=university *t*-score; VST=vocabulary size test score; HST-s=high school *t*-score; Ronin=number of years students were ronin; 8a, 8b, 8d, 9, 10, 11, 12, 14, 15=question number.

weak ( $\pm .30$ ), moderate ( $\pm .50$ ) and strong ( $\pm .70$ ) correlations.

A multiple-linear regression model was built using four items as the independent variables, high school *t*-score, I only read when I need to, my current university is my first choice, and my intended final educational attainment level, and the vocabulary size test score as the dependent variable. This was done to test the predictability of the learner capital questionnaire. The results of the regression indicated the four predictors explained 70% of the variance ( $R^2 = .49$ ,  $F(4,45) = 10.812$ ,  $p = .000$ ). For parsimony, items were removed backwards. A second regression indicated that two predictors, my current university is my first choice, and my intended final educational attainment level, explained 49% of the variance ( $R^2 = .24$ ,  $F(2,47) = 16.291$ ,  $p = .000$ ). It was found that first choice was a significant predictor ( $\beta = .419$ ,  $p = .001$ ), as was final educational attainment ( $\beta = .389$ ,  $p = .00$ ). Thus, certain items from the learner capital questionnaire were found to predict vocabulary size. Concurrent validity of the learner capital questionnaire would show that people high in learner capital are at higher-ranked universities and people low

should be at lower-ranked ones. For example, in this small sample discussed here, there were weak to strong correlations between current university and reading, university choice (self and parents) and future educational attainments. Furthermore, there were similar correlations between breadth of L2 vocabulary and reading, university choice and future educational attainments. On these and many of the other items, the questionnaire succeeding in splitting the participants into two groups as was predicted *a priori*. Although I have yet to investigate convergent and divergent validity, and although my sample is small, I tentatively claim that the learner capital questionnaire has criterion-related validity.

To test the reliability of the learner capital questionnaire, cronbach's  $\alpha$  was calculated with a random subset of seven items. The following items were used: university, high school *t*-score, reading is necessary, my current university is my first choice, my parents wanted me to attend this university, my parents were eager for me to study, and intended final educational attainment. Cronbach's  $\alpha$  was .521 indicating poor reliability.

## **Conclusion and Future Work**

The immediate goal of this paper was to provide further rationale for the learner capital questionnaire. First, different types of validity were discussed with examples given; and then these were applied to the learner capital questionnaire. The results, which showed that many of the items separate the participants into two groups of higher and lower learner capital, as predicted, allow me to cautiously claim that items of the learner capital questionnaire that were reviewed in this paper appear to reveal both translation validity and criterion-related validity. Another goal of the paper is attempt to answer the following question: How reliable are responses? Only a small set of questions were used with two small populations that were different from the outset.

Cronbach's  $\alpha$  was found to be poor. A third goal of this paper is identify preliminary results from certain sections of the questionnaire. These were described above. On many items and sections of this short sample, the two different populations were found to be different.

Going forward much work remains.

### References

- Apple, M. T. (2011). *The big five personality traits and foreign language speaking confidence among Japanese EFL students* (Unpublished doctoral dissertation). Temple University, Philadelphia, PA.
- Beglar, D. (2010). A Rasch-based validation of the vocabulary size test. *Language Testing*, 27, 101–118. doi: 10.1177/0265532209340194
- Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.) *Handbook of theory and research for the sociology of education* (pp. 241–258). New York, NY: Greenwood.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Coste, J., Guillemin, F., Pouchot, J., & Fermanian, J. (1997). Methodological approaches to shortening composite measurement scales. *Journal of Clinical Epidemiology*, 50(3), 247–252.
- Cronbach, L. (1990). *Essentials of psychological testing*. New York, NY: Harper & Row.
- Elwood, J. A. (2011). *Enriching structural models of L2 willingness to communicate: The role of personality, ego permeability, and perceived distance* (Unpublished doctoral dissertation). Temple University, Philadelphia, PA.
- Field, A. (2009). *Discovering Statistics Using SPSS*. London, UK: Sage Publications.
- Gardner, R. C., & MacIntyre, P. D. (1993). On the measurement of affective variables in second language learning. *Language Learning*, 43, 157–194.
- Gay, L. (1987). *Educational research: Competencies for analysis and application*. Columbus, OH: Merrill.
- Kariya, T. (2010). From credential society to “learning capital”. In H. Ishida & D. H. Slater (Eds), *Social Class in Contemporary Japan* (pp. 87–113). Milton Park, Oxon, UK: Routledge.
- Nation, I. S. P., and Beglar, D. (2007). A vocabulary size test. *The language teacher*, 31, 9–17.
- Nunnally, J. (1978). *Psychometric Theory*. New York, NY: McGraw-Hill.
- Nunnally, J., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Richard, J.-P. J. (2013). Building a questionnaire to measure learner capital. *Essays and Studies*, 64, 185–215.
- Sakui, K., & Gaies, S. J. (1999). Investigating Japanese learners' beliefs about language learning. *System*, 27, 473–492.



- Smith, Jr., E. V. M. (2001). Evidence for the reliability of measures and the validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, 2, 281–311.
- Sugimoto, Y. (2010). *An introduction to Japanese society* (3rd ed.). Cambridge, UK: Cambridge University.
- Trochim, B. & Donnelly, J. (2008). *The research methods knowledge base*. Mason, OH: Cengage Learning.
- Yamamoto, Y. & Brinton, M. C. (2010). Cultural capital in East Asian educational systems: the case of Japan. *Sociology of Education*, 83(1), 67–83.

Keywords

learner capital, translation validity, criterion-related validity, reliability