March 2003 Articles

ASSESSING JAPAN'S INSTITUTIONAL ENTRANCE REQUIREMENTS.

Headnote

Gregory Poole graduated from Brown University with a degree in anthropology and has since received graduate degrees in both linguistics (University of Surrey) and Japanese Studies (University of Sheffield). He first went to Japan in 1986 as part of the Monbusho English Fellow (MEF) Program and has since held numerous positions as instructor, tutor, and lecturer at both Japanese and British universities, as well as various private language schools. Presently he is working on an ethnography of higher education in Japan as well as looking at language learners as ethnographers while pursuing a doctorate in social anthropology. Greg can be contacted at: gregory.poole@anthropology.oxford.ac.uk

1. Introduction

Entrance into institutions of tertiary education in Japan has been described as a hierarchical system of 'exam hell' by more than one critic (see, e.g., Cutts 1997; Yoneyama 1999). Families invest enormous resources, both in time and money, to cram for the tests. In some cases, if initially unsuccessful at entering the college of choice, students even devote a year (or more!) after high school to further prepare to sit the exam yet again. To what extent this 'hell' affects in the same way the entire college-bound student population is open to debate. Nonetheless, in a competitive atmosphere these tests for entrance into colleges are held with great import by students, parents, institutions, and the general public. Considering the authority these college exams hold in Japanese society, a commensurate assessment of the quality of the tests themselves seems lacking. Although they may have privately held beliefs, for numerous reasons (beyond the scope of this paper) the entrance test developers themselves, as well as the institutions where they are employed, are especially hesitant to offer publicly data that would objectively evaluate these numerous admissions exams.

This short paper will attempt to begin to address this shortcoming in research on Japanese entrance exams not by offering any hardcore statistical analyses, but rather by simply calling attention to the need for more professional responsibility and research at the institutional level.

Specifically, the L2 proficiency portion of the entrance exam at a small university of commerce in Tokyo will be considered. The main question that will be addressed is whether either a standardized or performance-based language assessment would be more appropriate as part of the institutional entrance exam.

2. Standardized & Performance Proficiency Testing: Norm-referenced vs. Criterion-referenced

Numerous books on constructing and using language tests have been written (see, e.g., Lado 1961; Valette 1977; Hughes 1989; Bachman 1990; Alderson, Clapham et al. 1995; Brown 1996). In most of these discussions the authors classify a measurement strategy as either norm-referenced (NRM) (i.e., standardized) or criterion-referenced (CRM) (i.e., domain-referenced or "performance" tests) (Fulcher, personal communication). Brown (1995) clearly delineates the differences of these two types by focusing on the categories of "test characteristics" and "logistical dimensions."

To elaborate briefly, NRM's are general tests intended to be used to classify students by percentile for measuring either aptitude or proficiency for admissions into or placement within a program. CRM's, on the other hand, are more specific, achievement or diagnostic tests intended to be used for motivating students by measuring to what percent they have achieved mastery of the taught/learned material. Griffee (1995, p. 21) points out how the type of interpretation that a NRM, such as the Test of English as a Foreign Language (TOEFL), offers is the relative performance of the students compared with that of all the others resulting in, ideally, a bell curve distribution. Contrastingly, the CRM, such as a locally produced achievement test, measures absolute performance that is compared only with the learning objective, hence a perfect score is theoretically obtainable by all students who have a mastery of the pre-specified material, or conversely, all students may fail the test.

Closely related to the distinction between NRM and CRM tests is a theme that is repeated in much of the language testing literature (see, e.g., Brown 1996, p. v). This is the concern that since program-level decisions and classroom-level decisions are two entirely different categories of testing, different assessment instruments must be implemented. Admissions and placement decisions are questions of proficiency and students are ideally spread out in a continuum for which a NRM is the test of choice. Language skills are tested generally and students can then be grouped accordingly into ability levels for decisions of either admission into a program or streaming into different classes within a program. Comparisons of average proficiency levels within a program, or across institutions on a state, national, or international scale, are other program-level concerns that are best addressed with a NRM.

In classroom-level decision-making, on the other hand, diagnostic or achievement assessment is most helpful. For this end, CRMs are most accurate in helping teachers (and administrators) to assess the strengths and weaknesses of individual students with regard to the curriculum goals, as well as checking progress and achievement within such a program. Using data from CRM testing, teachers can be better equipped to assign final grades for a course and administrators can make more informed decisions about curriculum changes or adjustments.

This concern with matching informational needs with the appropriate testing instrument is critical to "sound testing." "Bad or mediocre testing is common, yet most language professionals recognize that such practices are irresponsible and eventually lead to bad or mediocre decisions being made about their students' lives." (Brown 1996, p. vii).
First, a very brief review of the relevant literature surrounding second language proficiency testing will be undertaken. What theoretical and logistical concerns must be addressed by the administrators and developers of such a test? To contextualize the discussion, the milieu in which the Japanese college entrance exams take place must be considered, and specifically, the situation at this college in Tokyo will be described. Finally, based on these discussions, an argument will be made for implementing a standardized, norm-referenced EFL proficiency test as part of the entrance exam to this institution.

2.1 Theoretical & Conceptual Considerations

Beyond this basic division of NRM and CRM test types are other theoretical considerations that must be made when developing a language test. Skehan (1988) points out how theory must be the driving force behind any language test. He mentions (p.211) that the oft heard proposal of solving the problem of language testing by following around one's students and observing their communicative performance in a variety of situations is not only impractical, biased, unethical, and unreliable, but, above all, the data simply may not be representative. The underlying challenge of language testing, then, is to achieve representativeness. Deciding what and how to test requires recourse to a specific linguistic theory (Henning 1987). The theoretical buck stops with testing since "testers, by researching into the structure of language proficiency, are attacking fundamental problems in language teaching and applied linguistics." (Skehan 1988, p. 211).

Davies (1978, p. 149) identified such a tension between assessment and pedagogy when he spoke of a testing continuum that ranges from the analytic to the integrative. Oller's (1979) theory of Unitary Competence Hypothesis (UCH) purported that integrative tests would be the best measurement tool, conflicting with previous, analytic assumptions (see Lado 1961; Davies 1973) which assumed skills-by-levels matrices. The Communicative Language Teaching (CLT) paradigm for language testing holds that performance, rather than standardization, should be the goal of measurement. Real life tasks, face validity, authenticity, and performance were four important key words surrounding the CLT model for testing (Fulcher 2000). This communicative approach, a psycholinguistic-sociolinguistic "stage" in the history of language testing (Weir 1990), was what Morrow (1981) coined "the Promised Land." Recently Fulcher (2000) reiterated Davies' emphasis on the similarity between the analytic-integrative and reliability-validity spectrums and pointed out how this debate on performance vs. discrete-item testing had actually taken place decades before.

In fact, the competence/performance issue is a concern in linguistics that has long interested language-testing experts, ever since the distinction was originally proposed by Chomsky (1965). Linguistic competence (a person's knowledge of an L1 or L2) is, obviously, more

perfect than actual linguistic performance (real language use) even in 'native speakers'. When testing L2 learners, then, an assessment of competence can only be guessed at indirectly by estimating a student's performance. Even a successful test of the underlying psychological construct of competence such as the TOEFL "only provides an estimate of the student's performance, which is only a reflection of the underlying construct, or competence." (Brown 1996, p. 29)

If measuring language proficiency of learners for admissions into university, then, administrators must be aware that if the assessment is, by definition, not entirely accurate, at the very least high standards must be upheld to ensure fairness in the selection process. In language testing for admissions, being objective necessarily implies addressing the concerns of both validity and, even more important in the eyes of this author, reliability. Does a test actually measure what it claims to be measuring-- for example, in a college admissions exam, general L2 proficiency (i.e., Is the test "valid"?)? More importantly, does the language portion of an entrance exam give the same objective results every time it is administered (i.e., Is the test "reliable"?)? Understanding the inherent inaccuracy of testing to begin with, most standardized NRM test developers rigorously analyze the validity and reliability of the tools they develop to ensure that the measurement of learner performance is at least as statistically accurate as possible.

2.2 Logistical Concerns

In addition to these lofty concerns of theoretical and statistical ideals, harsh reality forces administrators and teachers to compromise in addressing practical and logistical issues. How to allocate limited resources is a constant concern for any institution, especially universities. It is rarely true that infinite time and money exists to both teach and test languages, or any other subject for that matter. Test cost, construction, administration, and scoring are logistical concerns that all involve allocation of time and money. Most experienced administrators will admit (e.g., Brown 1996, p. 33) that language testing decisions are, more often than not, determined more by these practical concerns than by theories of language learning and assessment. Nevertheless, within the time/cost parameters that exist in a given entrance test environment, test developers and examiners have a professional responsibility to properly address issues of theory, validity, and reliability.

3. English Entrance Exams in Japan: cultural/situational considerations

Japan has one of the highest rates of post-secondary school attendance among all industrialized nations, with 2.5 million undergraduates enrolled at over 600 national, public, and private four-year universities (Hirowatari 2000). Over half of all Japanese teenagers, then, apply to take a college entrance exam for admission into a tertiary institution. Most such admissions exams include a compulsory English proficiency sub-test although EFL is not a state-required subject at primary, secondary, or tertiary schools in Japan. Partly because of this university entrance exam focus on English, while only a handful of students are exposed to language classes in primary school, over 10 million 12 to 18 year olds, and another million or so university

students, 'elect' to study English.

3.1 "Exam hell": *Hensachi*

In Japan there is a commonly held belief in "the educationally credentialized society," or *gakureki shakai*. In many cases, the extraordinary emphasis on ranking colleges and universities has led to a brand-name sensitivity that may affect a person for their entire life. One effect of a *gakureki shakai* is a phenomenon that has been labeled "exam hell."

As was mentioned in the introduction, the so-called "exam hell" is pressure felt by many young adults in Japan (as well as South Korea, Taiwan, Singapore, and other Asian countries for that matter). Most teenagers are expected to prove their intellectual mettle (or exam-taking talent) on these fact-oriented exams, even though they are rarely pushed to excel once they have matriculated at a college or university (see McVeigh 1997). Entrance into a university is often equated with passing the test, and in actuality this is often the case. Though admissions procedures are becoming more creative in recent years, the majority of colleges have resisted any change in a system that has been in place, arguably, since the Meiji Era in the late 1800s (Amano 1990). Indeed the university entrance, and overall education, system itself is inherently immobile (Frost 1991; Schoppa 1991), and has been described as a societal 'filtering' mechanism to create a class structure where otherwise none purportedly exists (see, eg, Cutts 1997; McVeigh 1997; McVeigh 1998).

Students are strictly ranked according to *hensachi*, the "abstract notion of a national norm-referenced person-indexed score." (Brown 1995, p.25). Using this score, high school and prep school teachers advise their students about which university entrance exams they should take based on their probability of acceptance (a high school teacher's reputation is on the line if their students shoot too high and miss their mark-- conservatism that is a necessity). In fact, the largest cram school syndicates in the Tokyo and Osaka area publish *hensachi* ranking lists of two and four-year colleges which students and teachers use to make application decisions.

3.2 Changing demographics

Japanese society is now faced with two demographic challenges that have been termed *shoushika* (low birthrate syndrome) and *koureika* (aging syndrome). These changes, of course, have repercussions throughout society, and schools are already witnessing the effects. Most universities in Japan have seen, first, a slowing in the rising rate of applicants, and, now, an overall decrease in the number of students sitting the yearly exams. This has forced a normally conservative sector of society to move in relatively innovative ways in an attempt to counteract their growing inability to attract students. Even top name schools in the higher echelons of the rankings have had to consider the ramifications of less and less applicants each year. Not least of their concerns is financial, of course, since entrance exam fees are a substantial source of revenue (in the $ millions) even for the prestigious, but inexpensive, national universities ($150 per student). No school in Japan can afford to sit on the laurels of past achievement and national prestige, least of all the colleges occupying the lower rankings. Recently college prep

and cram schools have instituted a new "F" rank, designating those colleges where the entrance exam is a mere formality since any student that applies is automatically accepted, given a "free pass."

3.3 Analyses of Japanese university entrance exams

With that being said, the conservative world of university entrance exams is slow to change. Though the doors are slowly opening since schools are acknowledging that they must lower the bar to keep the freshman class (and, ergo, the coffers) filled, the content of entrance exams, especially the English tests, have changed little. Although the administrators and board of directors may be anxious for progressive change, the test developers themselves control, for the most, the content of the entrance exams.

In fact, Brown & Yamashita (1995; 1995) conducted a longitudinal study of the content of English language entrance exams at Japanese universities and found insignificant change over a period of years. This would tend to indicate that university test developers are content to examine the students as always. After analyzing the content of 21 English entrance exams, Brown & Yamashita (1995; 1995) made many important findings, of which the following are most pertinent to this discussion:

1. Though the Japanese Ministry of Education guidelines implemented in 1993 strongly urged greater emphasis on listening skills, much of the content of the entrance exams still involve discrete-point grammar and translation items.
2. Considering that most teachers use simplified texts for classroom pedagogy, the readability indexes indicated that the text passages in the exams were quite difficult, ranging from a 'native'-speaker eighth grade (junior high) level to thirteenth grade (university) level.
3. The considerable variety of test items on an English entrance exam indicates that the ability to take exams may be more important than actual language proficiency.

It would appear to this author, that while the Ministry of Education has been encouraging a more communicative approach to language education, most university examiners must be either:
1. unsympathetic to the aims of the state education planners or
2. ignorant of proper testing methodology or
3. unaware of the negative effect of grammar-translation testing "washback."
Most probably, all three reasons are partly to blame for the present state of affairs in entrance testing. First, Marshall (1994) adeptly delineates the processes by which university professors have historically tended to buck bureaucratic *shido* ("guidance"). Secondly, Terauchi (2000) summarizes a Japan Association of College English Teachers (JACET) study that revealed an appalling lack of graduate training in applied linguistics and English Language Teaching (ELT) at university and college foreign language faculties. Finally, Mulvey (1999) has pointed out that even the possible positive effects of "washback" from testing grammar-translation, i.e., greater reading ability, have not occurred in Japan.

4. Edo College of Commerce Entrance Exams

Edo College of Commerce (ECC) is a pseudonym for a four-year business university in Tokyo with an undergraduate enrollment of 2800, as well as a very small graduate program. There is at the moment only one faculty-- commerce-- with two majors-- commerce and management. In 2001 the school expanded by adding a faculty of management with four new majors. The overall student enrollment has not been affected by this curriculum change, however. The school has a solid reputation as a school of business and commerce, but it is clearly a third, or fourth, tier institution. Although it has not yet been classified as an "F" rank school by the prep and cram school syndicates, students with high *hensachi* do not normally apply to or enroll at this school.

4.1 Entrance pathways at ECC

There are a total of nine different "pathways" for undergraduates students to enter ECC: 3 different "Recommendation" interview paths, a newly established "AO" (i.e., "Admissions Office") interview path for students with unique qualifications, a "Sports" scholarship pathway, a "Center" exam that is developed and administered by a central agency (much like the SAT in the U.S.), and, the "Regular" path by which three-quarters of the students enter ECC, two different locally-developed entrance exams administered at the beginning and end of February (n.b., the Japanese school year is from April to March). These "Regular" exams administered by faculty members of ECC are entirely developed by a team of ECC full-time professors of *kokugo* (Japanese), history, math, business, and foreign languages.

4.2 English proficiency testing at ECC

The English portion of these exams is developed from scratch in-house every year by the entire staff of six foreign language teachers. The content of the language exam is for the most part determined practically. Individual teachers are very conscious of completing the onerous task of test writing as swiftly and painlessly as possible. For this reason, the previous years exams are used as models and only slight modifications are made year-to-year. It is no surprise that since there is no resident testing "expert" (among the six faculty, only one holds a PhD in ELT, and less than half have masters level training in applied linguistics) and nothing but very rudimentary knowledge of language assessment, the full-time professors among the staff at ECC hold very little discussion regarding either appropriate testing methodology or exam quality. There is, however, much discussion and concern for "saving face," i.e., avoiding mistakes in clarity, grammar, spelling or misprints. There is also a lot of care taken with maintaining the security of the tests; they are kept under lock and key as they go through the various proof-reading processes over the course of three months (October- December).

Interestingly, the English exam has always been a required sub-test but became an elected test in 2001. The top-down decision to make this portion of the exam elective was made by the Admissions Committee seemingly in response to the falling number of applicants mentioned above. The reasoning is that by not requiring English more prospective students will be likely

to sit the entrance test, given the general aversion to English study among high school students in general and especially among the ECC applicants.

5. Conclusion: Rationale for Adopting a Standardized, Norm-referenced Entrance Exam at ECC

It is common knowledge in the fields of applied linguistics and ELT, and education in general, that there is no "perfect" test for any given purpose. What is necessary, however, is constant monitoring and improvement of any given testing program in order to ensure both language teachers and learners that the inevitable compromises are both conscious and rational decisions, not randomly based on 'blind' tradition. This current study shows that there is clearly room for improvement in language proficiency testing on the ECC entrance exam. Especially given the high face validity of all university entrance exams in Japan, or any other society, the test should be a rigorously analyzed instrument. "Evidence of validity should be presented for the major types of inferences for which the use of a test is recommended. A rationale should be provided to support the particular mix of evidence presented for the intended uses… For each total score… estimates of relevant reliabilities and standard errors of measurement should be provided in adequate detail to enable the test user to judge whether scores are sufficiently accurate for the intended use of the test." (American Psychological Association., et al. 1985, pp. 13-20)

In conclusion, after careful consideration of both the abovementioned theoretical and logistical concerns of both standardized and performance testing, as well as the state of affairs vis a vis entrance exams in Japan in general and at ECC specifically, rationale for adopting a norm-referenced language proficiency exam will be discussed. The arguments for choosing a NRM tool rather than a more performance-oriented CRM test will center around five issues of authenticity, measuring proficiency and not achievement, washback, practicality, and test evaluation and adoption.

The argument that performance-based tests are more 'authentic' than NRM tools is tenuous. By definition a language test is unnatural, no matter how 'communicative' the design and intent. "Any language test is by its very nature inauthentic, abnormal language behavior, for the test taker is being asked not to answer a question giving information but to display knowledge or skill."(Spolsky 1985, p. 39) This being said, standardized test developers are aware of the need for as much 'authenticity' as possible, witnessed by the TOEFL 2000 project, for example. Simply because a test is performance-based does not necessarily guarantee authenticity.

Furthermore, communicative, or performance, language tests (such as the CUEFL project) are characteristically criterion-referenced (Morrow 1991, p.112)-- excellent for diagnostic measurements but poor indicators of general language proficiency of learners in relation to other learners. If an institution is to make informed entrance decisions being fair and accountable for its admissions decisions, a norm-referenced test is essential.

Test 'washback' is often cited as an argument against standardized testing. Most informed test developers are conscious of the possibility of an exam's influence on language pedagogy. This

phenomenon, while not to be dismissed, is not very well understood even by experts. In fact, the complexities of washback, though implicating Japanese exams in a negative phenomenon, have not demonstrably shown with empirical evidence the alleged connection of NRM testing with outdated classroom pedagogy. Rather, could 'tradition' (or, more critically, lack of proper faculty development-- 'FD' is a popular catchphrase at Japanese universities recently) be the 'culprit' in a CLT gap in theory and practice in the Japanese EFL classroom? Nevertheless, if washback does exist in the entrance exam environment of Japan, it is certainly not because of 'standardized' testing, a term that is an oxymoron in the world of Japanese university admissions. Must not educators provide a more 'standard' assessment tool with which to decide the future lives of millions of Japanese youth?

At ECC, as at nearly every college in Japan, an admissions office, or even secretary, does not exist. There is no such title, or official role, of "admissions counselor." The burden of admissions policies, college recruiting, test development, and actual administration of these exams is all carried out by a rotating team of faculty and administrative staff, further supported by ad hoc committees. None of the teachers or administrative staff involved was hired specifically for any of these admissions related jobs. Those 'unlucky' faculty members and staff chosen for administrating the admissions process do an admirable job, given the inherent limitations and incredible logistical complexities of handling, and testing, thousands of applicants yearly. Quite simply, a performance type of entrance exam would be totally unrealistic (and probably vetoed by the entire faculty!) given the inevitable logistical complications of testing and marking an open-ended, interactive language exam.

Finally, an in-house entrance testing program developed, modified, and ostensibly improved, yearly to meet the specific needs of the ECC applicant population and language program would be ideal. Such a project, however, is problematic (if not impossible) given the (wo)man-hour resources and expertise available for the task. There is a clear need for considering the adoption of a commercially available, standardized NRM as a tool for making appropriate admissions decisions. Though not ideal, such a test would be a vast improvement over the unscientific testing approach being implemented at present. Such a test would have the added benefit of situating ECC applicants in a national and international macrocosm, aiding school executives, administrators, and faculty to make more informed intelligent admissions, language program, and classroom level decisions.

6. References

Alderson, J. C., C. Clapham, et al. (1995). *Language test construction and evaluation. Cambridge, Cambridge University Press.*
Amano, I. (1990). *Education and examination in modern Japan.* Tokyo, Tokyo University Press.
American Psychological Association., American Educational Research Association., et al. (1985). *Standards for educational and psychological testing.* Washington, DC, American

Psychological Association.

Bachman, L. F. (1990). *Fundamental considerations in language testing. Oxford [England] ; New York, Oxford University Press.*

Brown, J. D. (1995). Differences between norm-referenced and criterion-referenced tests. *Language Testing in Japan.* J. D. Brown and S. O. Yamashita. Tokyo, The Japan Association for Language Teaching: 12-19.

Brown, J. D. (1995). *"A gaijin teacher's guide to the vocabulary of entrance exams." The Langage Teacher(December):* 25.

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

Brown, J. D. and S. O. Yamashita (1995). "*English language entrance examinations at Japanese universities: what do we know about them?" JALT Journal* 17(1): 7-30.

Brown, J. D. and S. O. Yamashita, Eds. (1995). *Language Testing in Japan.* JALT Applied Materials. Tokyo, The Japan Association for Language Teaching.

Chomsky, N. (1965). *Aspects of the theory of syntax. Cambridge*, M.I.T. Press.

Cutts, R. L. (1997). *An empire of schools : Japan's universities and the molding of a national power elite*. Armonk, N.Y., M.E. Sharpe.

Davies, A. (1973). *The construction of language tests. The Edinburgh course in applied linguistics: vol.4. testing and experimental methods. J. P. B. A. a. A. Davies. London, Oxford University Press.*

Davies, A. (1978). "*Language testing: survey article 1." Language teaching and linguistics abstracts* 11: 145-159.

Frost, P. (1991). Examination hell. *Windows on Japanese education.* E. R. Beauchamp. New York, Greenwood Press: 291-305.

Fulcher, G. (2000). "*The 'communicative' legacy in language testing." System 28(4)*: 1-15.

Griffee, D. T. (1995). Criterion-referenced test construction and evaluation. *Language Testing in Japan*. J. D. Brown and S. O. Yamashita. Tokyo, The Japan Association for Language Teaching: 20-28.

Henning, G. (1987). *A guide to language testing : development, evaluation, research.* Cambridge, Mass., Newberry House Publishers.

Hirowatari, S. (2000). "*Japan's national universities and dokuritsu gyosei hojin-ka." Social Science Japan 19(September)*: 3-7.

Hughes, A. (1989). *Testing for language teachers. Cambridge [Cambridgeshire] ; New York, Cambridge University Press.*

Lado, R. (1961). *Language testing : the construction and use of foreign language tests ; a teacher's book. London, Longman.*

Marshall, B. K. (1994). *Learning to be modern : Japanese political discourse on education.* Boulder, Westview Press.

McVeigh, B. J. (1997). Life in a Japanese women's college : learning to be ladylike. London ; New York, Routledge.

McVeigh, B. J. (1998). *The nature of the Japanese state : rationality and rituality. London ; New York, Routledge*.

Morrow, K. (1981). Communicative language testing: revolution or evolution? *The*

communicative approach to language teaching. C. J. Brumfit and K. Johnson. Oxford, Oxford University Press: 143-157.

Morrow, K. (1991). Evaluating communicative tests. *Current developments in language testing*. A. Sarinee and Regional Language Centre. Singapore, SEAMEO Regional Language Centre: 111-118.

Mulvey, B. (1999). "*A myth of influence: Japanese university entrance exams and their effect on junior and senior high school reading pedagogy." JALT Journal 21(1)*: 125-142.

Oller, J. W. (1979). *Language tests at school : a pragmatic approach*. London, Longman.

Schoppa, L. J. (1991). *Education reform in Japan : a case of immobilist politics. London ; New York, Routledge*.

Skehan, P. (1988). "*Language testing. part 1." Language Testing 21*: 211-221.

Spolsky, B. (1985). "*The limits of authenticity in language testing." Language Testing 2*: 31-40.

Terauchi, H. (2000). *English for Academic Legal Purposes in Japan*. Tokyo, Liber Press.

Valette, R. M. (1977). *Modern language testing*. New York, Harcourt Brace Jovanovich.

Weir, C. J. (1990). *Communicative language testing*. New York, Prentice Hall.

Yoneyama, S. (1999). *The Japanese high school : silence and resistance*. London, Routledge.