
An Evaluative Commentary of the Grade 1 EIKEN Test

GABRIELLE PIGGIN

Ritsumeikan Asia Pacific University, Japan

Bio Data:

Gabrielle Piggin is a Junior Lecturer at the Centre for Language Education, Ritsumeikan Asia Pacific University. She has been a language teaching practitioner for the past nine years both in New Zealand and Japan. Her research interests include Critical Language Testing, Critical Discourse Analysis, Sociolinguistics; gender and identity and second language acquisition.

Abstract

This paper aims to analyse the fit between the “use” and “usefulness” of the EIKEN Grade 1 test as a valid and reliable instrument to measure second language proficiency. The recent re-positioning of the EIKEN Grade 1 test as an internationally recognised higher stakes test of English proficiency, which allows successful test taker’s access to higher education in international academic contexts, is discussed. The evaluation was based upon Bachman and Palmer’s (1996) model of test usefulness in order to critically ascertain whether such test use is one that could be justified. This empirical investigation explicitly purports the need for further quantitative and qualitative analysis to justify EIKEN Grade 1’s test usefulness.

Introduction

This paper evaluates Japan’s Grade 1 EIKEN test in regards to Bachman and Palmer’s (1996) seminal model of test usefulness. EIKEN is recognised as the most popular domestic public test of English language proficiency in Japan, and in recent years the Grade 1 level has ‘gained’ a new purpose; validating Japanese students entry into graduate and postgraduate programmes of over 300 international educational institutions (“EIKEN: Recognition,”2010). Despite the dominant use of EIKEN in Japan’s language testing context there is little evidence which can be

looked to, to support Grade 1's usefulness as an instrument to measure the English language proficiency of testees entering international educational programmes, or the claims of Grade 1's equivalency to other large scale public language tests that EIKEN purports ("EIKEN: Grades," 2010). Additionally, comprehensive test specifications are unavailable in English ("EIKEN," 2010) thus; any evaluation of Grade 1 will remain somewhat ambiguous without undertaking considerable qualitative and quantitative research (Bachman & Palmer, 1996).

To understand exactly what further research is required to identify and evaluate a contemporary suitable balance of test qualities for this particular language test (see Appendix 3), it is anticipated that by an informal appraisal of the usefulness, in light of the limited evidence available, this paper will attempt to highlight the problematic issues related to EIKEN Grade 1, its test tasks and their subsequent use as an overall measurement of testees' language ability in relation to the target language use (TLU) domain of Western educational contexts (Bachman & Palmer, 1996). Bachman and Palmer's (1996) model considers the intersection of six test qualities; construct validity, reliability, authenticity, interactiveness, impact, and practicality which "all contribute to test usefulness, so that they cannot be evaluated independently of each other" (Bachman & Palmer, 1996, p. 34). To undertake this evaluation from an informed perspective; relevant language testing literature, the Grade 1 2009_2 Stage 1 test itself, ("EIKEN: Downloads," 2010) EIKEN's official English website, ("EIKEN," 2010) and personal communication with a former- Grade 1 testee (cited as A* to retain their anonymity) from whom anecdotal evidence has

been collected to present some 'real life' interpretations of testee's views of the test (see Appendix 7).

EIKEN: Overview

EIKEN, "Test in Practical English Proficiency", established in 1963 and developed by the non-profit organisation, the Society for Testing English Proficiency (STEP), was officially accredited by Japan's Ministry of Education (MEXT) in 2000 (Gottlieb, 2005, p. 33). STEP presently produces and administers EIKEN as "a seven-level suite of tests" that range from 'beginner' level Grade 5, to 'advanced' level Grade 1 ("EIKEN," 2010). Recently, Grade 1, and the bridging level Grade Pre-1, are "[r]ecognized for international admissions to graduate and undergraduate programmes" ("EIKEN: Recognition," 2010). The promotion for STEP's presentation at the 2007 JALT conference sums up EIKEN's entry into the academic entry sector of language tests.

Until four years ago, EIKEN was virtually unheard of outside Japan. Now the test is recognized for admissions at hundreds of institutions in North America and Australia. What happened? The growth of EIKEN from domestic test to global qualification ("eltcalender," 2007).

Grade 1 (see Appendix 1) is administered in two stages; "a paper-and-pencil test that evaluates reading, listening, and writing ability (...) a speaking test in the format of a face-to-face interview with two examiners" ("EIKEN: Overview," 2010). Grade 1 is a norm-referenced test and 70% is the cut-off score for both stages; the criteria presented in the 'EIKEN Can-do' document (see Appendix 2) which was compiled from a long term qualitative study (Yanase&Matsudaira, 2007; Dunlea,

2010) of what successful testees “believe they can accomplish in English in real-life situations” is not taken into account in scoring (“EIKEN: Research,” 2010).

EIKEN has three testing periods a year, at over 400 “official test centers” (“EIKEN: Administration,” 2010). Grade 1 testees are permitted to re-sit their Stage 2 interview up to three times, as long as they pay the testing fee of 7,500 yen (US 90 dollars) each time (“EIKEN: Overview,” 2010). Hence, Grade 1 is an accessible, affordable option, which is relatively more authentic and interactive than other domestic tests, and by allowing room for second and third chances in the anxiety-inducing oral interview (see Appendix 7) it is the preferred testing option for Japanese testees. While practicality is more than sufficient, the question remains; “How useful is this particular test for its intended purpose?” (Bachman & Palmer, 1996, p. 17).

Grade 1’s Usefulness

The main concern of EIKEN Grade 1’s contingent use as a high stake’s measurement of English proficiency for entry into international educational institutions is the lack of justification that supports this purpose therefore, the construct validity of the test is questionable, hence its overall usefulness (Bachman & Palmer, 1996). STEP, as a testing organisation, does not define nor refer to the construct of ‘language proficiency’ it seeks to measure definitively apart from the claim that the EIKEN levels cover “a broad spectrum of language ability” (“EIKEN: Research,” 2010). STEP also does not publically provide detailed test specifications in English, that is an “official statement about what the test tests and how it tests it” despite the English website being the only means of access to, and a means of recruitment for

international institutions who wish to 'recognise' EIKEN (Alderson, Clapham, & Wall, 1995, p. 9). For example, the Speaking interview's scoring scale which may contribute to a more effective affirmation of construct validity and reliability of the test tasks involved is unavailable (Bachman & Palmer, 1996; Hughes, 2003).

In comparison to the significant amount of research and critique that other major testing bodies such as IELTS, TOEFL®iBT, and the recent Pearson Test of English (PTE) Academic, have provided, been subject to, and subsequently applied, further substantial research and critique is necessary to validate EIKEN Grade 1's use as a high-stakes international measurement of proficiency (Fulcher, 2010). Currently, research related to justifying Grade 1's use only includes twelve 'in-house' studies which are available on the website under the sub-heading "Demonstrating validity: a list of recent projects" ("EIKEN: Research," 2010). Dunlea and Matsudaira (2009) are in the process of a long term project which involves series of small-scale (13 non-testing professional participants) standard setting workshops that are attempting to correlate EIKEN's higher grades with the levels of the well-established Common European Framework (CEFR). Their recent work is the basis for the equivalency claims which are presented on the website of Grade 1's correlation with CEFR level C1 (EIKEN, 2010). Brown's brief case study of EIKEN in his 2008 article concerning 'Texting Context Analysis' examines how his suggested framework can be applied to help "justify the use of language assessment to stakeholder's" (Brown, 2008, p. 275). Additionally, Brown's 'in-progress' research comparing EIKEN to TOEFL®iBT is anticipated by STEP as validating the usefulness of the EIKEN test in international testing literature ("EIKEN: Research," 2010).

Nielsen's (2000) study (which is not cited by STEP) is the only attempt to date to investigate EIKEN's construct validity and reliability quantitatively.

In addition to EIKEN's conspicuous absence on international testing websites (Fulcher, 2010), the only 'comprehensive' source of information available is EIKEN's English public website in English which was created in March, 2010 by STEP ("EIKEN: History," 2010). The website could be considered to present a biased perspective as most of the information available is geared towards the marketability and practicality of the EIKEN test instead of a critical, combined evaluation of the test qualities to situate its' usefulness ("EIKEN: History," 2010).

In this ever-increasing global world, language qualifications are an increasingly sound investment for second language learners to make, with the added attraction of a gateway to overseas study and qualifications it is undoubtedly an even sounder investment; in this respect, the face validity of Grade 1 is more than adequate (Hughes, 2003). However, Bachman and Palmer (1996) assert "the most important quality of a test is its usefulness" (p. 17). They emphasise that all of the six test qualities of their model "contribute in unique but interrelated ways to the overall usefulness of a given test" (Bachman & Palmer, 1996, p. 18) and the congruous tension of the six qualities "must be determined for each specific testing situation" (Bachman & Palmer, 1996, p. 135) (see Appendix 3). Inexplicably, STEP claims that its test developers refer to this model "to help make transparent principled decisions regarding the optimal balance of various features for each grade" ("EIKEN: Research," 2010) yet does not clarify how this has been achieved. However, the limitations of the past five decades of its' test revision is

acknowledged and STEP asserts that the two upcoming studies (Brown; Dunlea&Matsudaira) allow sufficient confidence in “offering these tests as robust measurement tools relevant to and appropriate for the needs and requirements of each grade” (EIKEN: Research,” 2010). Regardless of the limited evidence provided and unsubstantiated claims made by STEP, Grade 1 is attaining a wider global platform (“EIKEN: Recognition,” 2010).

Construct Validity

As Grade 1 is a high stakes test, STEP alleges that in its test development, validity and reliability are prioritised over practicality (“EIKEN: Research,” 2010). This may be evident in Grade 1’s Stage 2 which involves a face to face interview with two examiners thus increasing the reliability and validity of the subjective scoring (Hughes, 2003).

STEP admits that EIKEN’s test development “predate[s] much of the work on validity theory” (“EIKEN: Research,” 2010) which is problematic to interpreting test scores meaningfully and appropriately (Bachman & Palmer, 1996). Although STEP acknowledges that it has to “explicitly move beyond the strong content validation focus and reliance on a consensus-based approach which previously formed the core of justification for the uses and interpretations of EIKEN tests and test scores” it is yet to evidence how and when it intends to do this and if test tasks can be considered relatively more authentic as “content validity’ implies (“EIKEN: Research,” 2010). For instance, items are created specifically for every test which ensures test security however, no research has been undertaken to account for items ‘construct validity (“EIKEN: Research,” 2010). For example, Reading 1 section is not

in fact a direct reading test; rather vocabulary recognition in a short context which Hughes (2003) describes as “recognising appropriate word for context” (p. 182). Both reading sections and listening sections are composed of multi-choice questions (MCQ). Hughes asserts that MCQ are not appropriate in advanced tests, and despite their reliability as objective scoring methods there are many problematic issues concerning their validity, authenticity and interactiveness (see Hughes’ p. 75-78).

Additionally, the construct validity of the Reading and Listening sections’ test tasks do not effectively correspond with the TLU tasks that testees may be required to perform when studying abroad as they are too limited in scope (Weir, 1997; Brindley, 1997). The limited range of skills in the Reading Section (see Appendix 5) and the Listening Section (see Appendix 6) which are tested can be considered an ambiguous interpretation of generalising “beyond the testing situation itself to a particular TLU domain” (Bachman & Palmer, 1996, p. 21). Despite the new purpose of Grade 1 which positions the test as a life changing opportunity on a global scale, rather than a solely a domestic recognition of ability, its construct validity continues to be based upon language knowledge at the expense of pragmatic and sociolinguistic competence as exemplified in the Reading section 1 (MCQ) (Bachman & Palmer, 1996). However, STEP attains that content validity has been “constantly reviewed and verified in an open and interactive manner with all stakeholders” (“EIKEN: Research,” 2010) due to the review of the tests items and the take home, public format of the tests.

Reliability

Hughes (2003) asserts, “[t]he more important the decisions, the greater the reliability we must demand” (p. 39). As Grade 1 is a high stakes test, Bachman and Palmer (1996) suggest that logically “the test developer would want to set the minimum acceptable level [of reliability] very high” (p. 135). It is a cause for concern for all stakeholders that thorough quantitative research has not yet been undertaken by STEP to evaluate “an appropriate statistical estimate of reliability” (Bachman & Palmer, 1996, p. 135). Nielsen (2000) points out that although the multi-choice question format which is used for the reading and listening sections ensures “greater reliability” in the form of objective scoring, his small scale study concluded that the reliability co-efficient of 0.86 is considered quite low for a high-stakes test. Therefore, it is imperative for STEP to provide evidence of the consistency of measurement based upon characteristics of the test tasks to support their claims of reliability (Bachman & Palmer, 1996).

Authenticity and Interactiveness

In considering the intersection of the test tasks and their relation to the TLU domain, STEP claims that the items of each grade are tailored in their topical content “to be more relevant to typical test takers for that grade” (“EIKEN: Grades,” 2010). However, there seems to be a general lack of correspondence with the TLU domain of Western educational settings as both the Listening and Reading sections are not overtly reflective of academic ‘real life’ language, but of the professional world of work (see Appendix 4).

Despite STEP's claims that the content of the EIKEN tests is made "as relevant as possible to the learners" ("EIKEN: Research," 2010) Grade 1's 'relevance' is difficult to ascertain as a clear set of the characteristics of; Grade 1's testees, TLU tasks at the educational institutions which recognise EIKEN, and the Grade 1 test tasks are necessary to "determine the relative authenticity or interactiveness" of the test (Bachman & Palmer, 1996, p. 28). How useful Grade 1's test tasks are to measure testees' language ability relative to the TLU domain remains a grey area which future research needs to investigate while considering the sets of characteristics involved (Bachman & Palmer, 1996).

Impact

Shohamy (cited in Lazaraton, 2010) has been a pioneer of directing research away from the statistical interpretations of tests with her "focus on (...) social and political implications" (p. 261). She asserts, "reliability and validity, so what? What will it's impact be?" (Lazaraton, 2010, p. 261). Messick (cited in McNamara & Roever, 2006, p. 12) also explicates the effective interaction between tests and the socio-cultural context in which they are developed and used; "our conceptions of what it is that we are measuring and the things we prioritize in measurement, will reflect values, which we can assume will be social and cultural in origin." Brown's (2008) study proffers a suitable analytical framework for STEP to examine "the relationships between the tests and the various stakeholders" (p. 276) and assess the micro and macro levels of impact in regards to Japanese society to gain a more accurate perspective on not only the impact, but also the authenticity and interactiveness of the test tasks.

STEP states that EIKEN tests have always endeavoured to “enhance positive washback and contribute to the improvement of language education in Japanese society”, however the social implications have not yet been investigated (“EIKEN: Research,” 2010). In Japanese society EIKEN can be considered as having negative washback at the micro- level of teaching and learning and a negative impact on wider society (Bachman & Palmer, 1996; Hughes, 2003). EIKEN reinforces that English is a ‘skill’ and language knowledge of lexical items which are memorised and rote learnt (which is the dominant teaching methodology in Japanese language education) will lead learners to ‘proficiency’ (Gottlieb, 2005). This is seen in the number of learners who are learning only for the test tasks; for example, when testees study, they use EIKEN produced learning materials (see Appendix 7).

Practicality

From the following statement it is clear that STEP does not lack resources; “It takes a team of 200 writers, editors, and content specialists more than four months to produce one set of EIKEN tests” (“EIKEN,” 2010). Therefore, issues of practicality should not hinder any future developments made to the test. Undeniably, EIKEN is an testing option catering for Japanese testees; Japanese rubric in the listening section, Japanese-only application forms, access to Speaking interview samples and EIKEN learning materials are only available from the Japanese website (“EIKEN: Products,” 2010). EIKEN’s accessibility through its wide distribution of test sites, frequency of testing periods, and sustained affordability, in addition to a fairly quick turn over of results, allows it to maintain its position as Japan’s most popular test (“EIKEN: Administration,” 2010).

From a Critical Language Testing view, Grade 1's marketability agendas outweigh any of the test qualities aside from practicality. The ethics of the "free publicity" and increase of Japanese students promised by STEP's "recognition form" for educational institutions who "sign up" is dubious ("EIKEN: Recognition," 2010). Certainly when high stakes decisions are based on a test which may be lacking in "minimal acceptable levels" of test usefulness, the consequences for the stakeholders involved are a pressing concern (Bachman & Palmer, 1996, p. 19).

Future Considerations

Grade 1's re-positioned purpose necessitates more evidence of the test's usefulness to justify the interpretation of the testees' scores and support STEP's claims (Bachman & Palmer, 1996, p. 135). For instance, if implications of future research elucidate the low construct validity and relatively less authentic 'Reading 1' test tasks for the TLU domain, instead of utilising resources for writing such test items), new more useful test tasks could be developed ("EIKEN: Research," 2010). In the past, concerns about EIKEN's reliability have been of issue (Nielsen, 2000) however, this discussion has revealed that the test qualities of construct validity, authenticity, interactiveness, and impact are of key concern in regards to Grade 1's usefulness to measure testees' ability to undertake academic study in an overseas context.

All language tests are inherently flawed as they are measuring intangible psychometric qualities however; efforts must be made to operationalise the best measures possible, especially for high-stake tests (Bachman & Palmer, 1996; Hughes, 2003; Lazaraton, 2010). From this empirical evaluative commentary it seems implicit that there is a gap between language testing theory and the 'practice' of the EIKEN

Grade 1 as a test for entry into educational institutions. Initially, Bachman and Palmer's (1996) checklist for "Developing a plan for the evaluation usefulness" would be an accessible starting point to further research (p. 133-156).

Conclusion

In sum, it seems that the issues that have been posited by this brief evaluation imply, like all testing issues, that there is no easy solution (Bachman & Palmer, 1996). STEP admits that in their development of EIKEN as a seven level framework, such "[f]lexibility (...) presents particular problems for validation" ("EIKEN: Research," 2010). Thus, a re-negotiation of the Grade 1 test as a separate entity rather than just another level of the EIKEN framework needs to be appraised in light of recent developments, theories, and debates concerning the theoretical construct of language ability itself and the usefulness of test tasks that endeavour to measure testees' language ability for 'real life' use (Bachman & Palmer, 1996; Hughes, 2003, Lazaraton, 2010). As Bachman and Palmer (1996) suggest further "procedures for collecting qualitative and quantitative evidence" need to be undertaken by both STEP and impartial language testing professionals to ensure that Grade 1 as a high-stakes test has the best feasible balance possible of the six test qualities (p. 134).

References

- Alderson, C., Clapham, C., & Wall, D. (1995). *Language test construction and validation*. Cambridge: Cambridge University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford:Oxford University Press.

- Brindley, G. (1997). Assessing listening abilities. In W. Grabe (Ed.), *Annual Review of Applied Linguistics* (pp. 171-191).
- Brown, J. D. (2008). Testing-context analysis: assessment is just another part of curriculum development. *Language Assessment Quarterly*, 5(4), 275-312.
- Dunlea, J. (2010). The EIKEN Can-do List: improving feedback for an English proficiency test in Japan. In L. Taylor & C. Weir (Eds.), *Language testing matters: Investigating the wider social and educational impact of assessment - Proceedings of the ALTE Cambridge Conference, April 2008*, *Studies in Language Testing* volume 31. Cambridge: Cambridge University Press.
- Dunlea, J., & Matsudaira, T. (2009). Investigating the relationship between the EIKEN tests and the CEFR. In N. Figueras & J. Noijons (Eds.), *Linking to the CEFR levels: Research perspectives*. Arnhem: CITO and EALTA.
- EIKEN. (2010). Retrieved from <http://www.stepeiken.org/>
- eltcalendar. (2007). Retrieved from <http://www.elcalendar.com/events/details/3664>
- Fulcher, G. (2010). Language testing resource website. Retrieved from <http://languagetesting.info/>
- Gottlieb, N. (2005). *Language and society in Japan*. Cambridge: Cambridge University Press.
- Hughes, A. (2003). *Testing for language teachers* (2nd Ed.). Cambridge: Cambridge University Press.
- Fulcher, G. (2010). Language testing resources. Retrieved from <http://languagetesting.info/>

- Lazaraton, A. (2010). From cloze to consequences and beyond: An interview with ElanaShohamy. *Language Assessment Quarterly*, 7(3), 255-279.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Oxford: Blackwell Publishing.
- Nielsen, B. (2000). Determining test reliability and quality of EIKEN test items: A statistical analysis of first year Kosen students responses to test items of an EIKEN third level test. *Research reports: Kushiro Technical College*, 34, 81-91.
- Weir, C. J. (1997). The testing of reading in a second language. In C. M. Clapham & D. Corson (Eds.), *Encyclopaedia of language and education* (pp. 39-49). Dordrecht: Kluwer Academic Publishers.
- Yanase, K., & Matsudaira, T. (2007). 英検Can-do調査. [The EIKEN Can-do Survey]. Paper presented at the 10th Japan Language Testing Association Annual Conference.

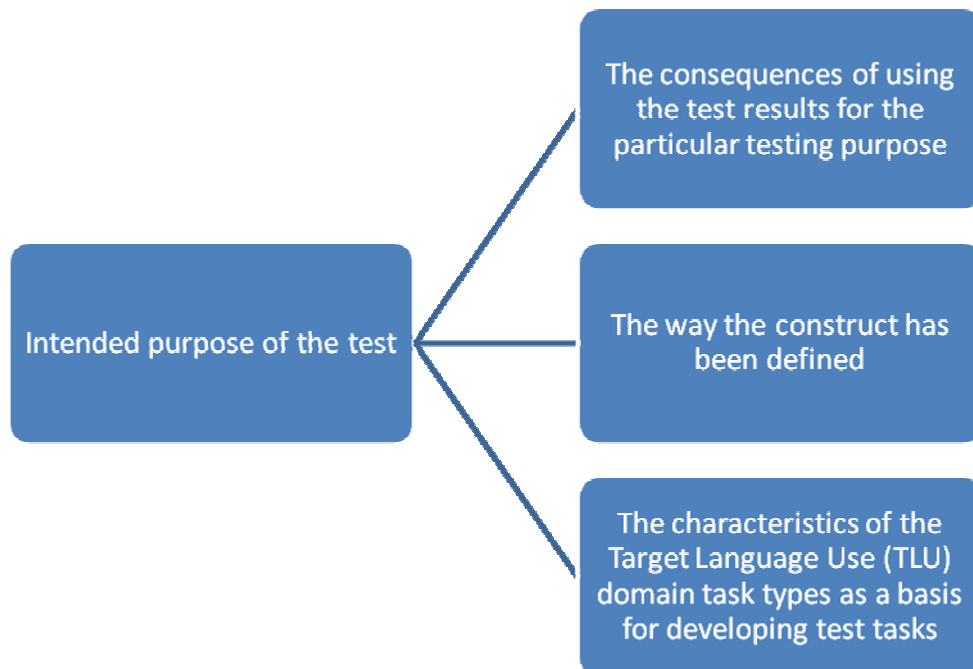
Appendix 1: Overview of Grade 1 2009_2 test ("EIKEN:Grades," 2010)							
Stages	Sections	Test tasks	Number of items	Weighting	Time Allocated	Section weighting	Scoring
Stage 1: Paper and Pencil test	Reading	1) Multiple-choice questions (MCQ). Recognise appropriate lexical item for short context texts.	25	25	1 hour 40 minutes	51 113 =45.13%	Objective-Quantitative
		2) Two texts with three MCQ each. Recognise appropriate phrase for passage.	6	6			
		3a) Two texts with three MCQ each. Recognise appropriate answer.	6	12			
		3b) One text with four MCQ. Recognise appropriate answer.	4	8			
	Writing	One 200 word 'English composition' about one topic including three sub-topics from the six available choices.	1	28		28 113 =24.78%	Subjective-N/A
	Listening	1) Ten short 'dialogue' listening texts. One MCQ for each text.	10	10	30 minutes	34 113 =30.09%	Objective - Quantitative
		2) Five monologues with two MCQ each text.	10	10			
		3) Five listening texts with one MCQ each text.	5	10			
		4) One long listening text with two MCQ.	2	4			
	Cut- off score to pass Stage 1 is 70%=79.1/113.						
Approximately one month later subject to passing Stage 1.							
Stage 2: Oral face to face interview with two examiners	Speaking	1) 'Free conversation' - answer examiners' questions	Not publically available in English	Not publically Available in English	1 minute	Not publically available in English	Subjective-N/A
		2) Preparation time for 'speech' about selected topic.			1 minute		
		3) Speech			2 minutes		

		4) Answer examiners' questions about speech topic .			4 minutes		
--	--	---	--	--	-----------	--	--

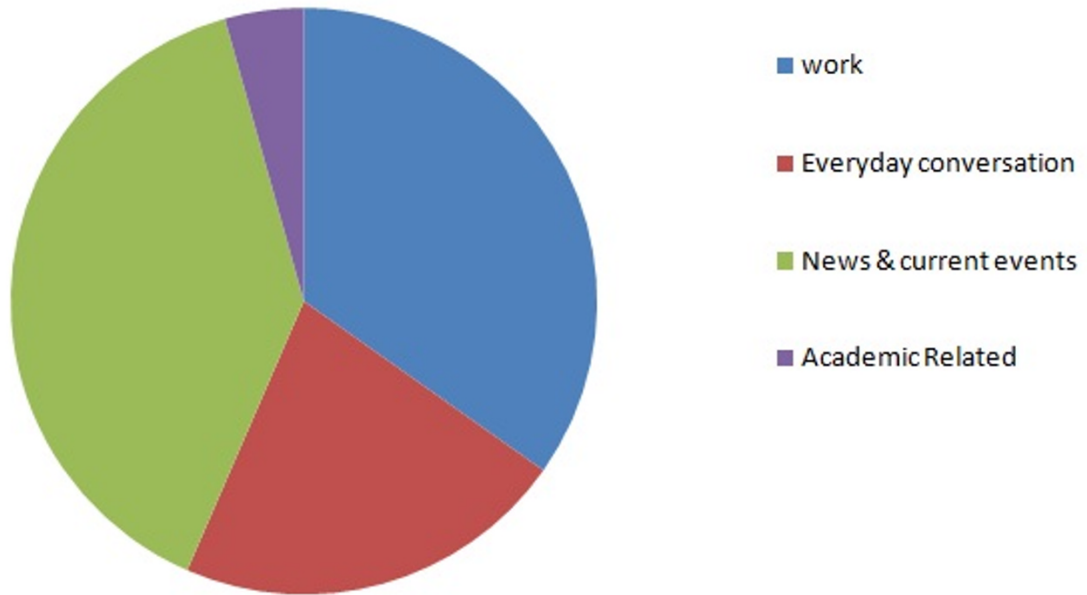
Appendix 2: Grade 1 'Can-Do' statements ("EIKEN Can-do list," 2010, p.6)

Reading	Can understand a wide variety of texts from a range of social, professional, and educational situations.
Writing	Can write at length about topics relevant to a range of social, professional, and educational situations.
Listening	Can understand a wide variety of content from a range of social, professional, and educational situations.
Speaking	Can take part in interaction on a wide variety of topics relevant to a range of social, professional, and educational situations.

Appendix 3: Considerations “to specify initial minimum acceptable levels for the qualities of usefulness” (Bachman & Palmer, 1996, p. 134)



Appendix 4: Overview of Grade 1 2009_2 Reading 1 test task topics



Appendix 5: "Operations in Reading" Grade 1 2009_2 Reading Section (Weir, 1997, p. 43)

Reading	1) Multiple-choice questions (MCQ). Recognise appropriate lexical item for short context texts.	Reading carefully for local comprehension
	2) Two texts with three MCQ each. Recognise appropriate phrase for passage.	Reading carefully for local comprehension
	3a) Two texts with three MCQ each. Recognise appropriate answer.	Reading carefully for global comprehension
	3b) One text with four MCQ. Recognise appropriate answer.	Reading carefully for global comprehension

**Appendix 6: "Listening Skills Taxonomy" of Grade 1 2009_2 Listening Section
(Brindley, 1997)**

Listening	1) Ten short 'dialogue' listening texts. One MCQ for each text.	Understanding meaning not explicitly stated.
	2) Five monologues with two MCQ each text.	Extracting spoken information from a spoken text.
	3) Five listening texts with one MCQ each text.	Extracting specific information from a spoken text.
	4) One long listening text with two MCQ.	Understanding meaning not explicitly stated.

Appendix 7: Anecdotal Evidence from A* concerning Grade 1 (A* Personal Communication, November, 8, 2010)**Construct Validity:**

- 1) "I don't think it shows my English ability because I prepared for an examination, I read the books for EIKEN, I understand the questions and I can pass. I don't think EIKEN can tell my English proficiency exactly"
- 2) "I think I prepared for the EIKEN questions very hard, that's why I passed the exam."
- 3) "It's too focussed on knowledge, like vocabulary"
- 4) "Even if I got the first grade, I don't think my English ability is so good"

Reliability:

- 1) "Actually I took several times...more than three times, always the same"

Authenticity/Interactiveness:

- 1) "Listening is very hard for me as I'm not accustomed to listening to speaking in Japan-I took traditional Japanese English lessons so when I was a high school student I had no ALT, I had no listening...just tapes...or the Japanese teacher said repeat after me. I had no opportunity to listen and speak, so it's very, very hard for me. I'm afraid of listening the most"
- 2) "Speaking is also difficult for me...I remember I was very, very nervous in front of the two interviewers. When I took the interview test I was sure I failed"
- 3) "Actually the interview was very unnatural-but better than TOEFL talking to computer screen it's very strange. In terms of speaking tests EIKEN is not so bad"
- 4) "So if it improves more it is best for Japan and Japanese, I think change vocabulary and writing longer and different speakers of listening, not just American"

Impact:

- 1) "But I think good impact, because students who are motivated can take a test easily"
- 2) "Only to study...of course I want to get qualification but my students asked me... "Did you get Grade 1?" several times so I am determined to get EIKEN Grade 1"
- 3) "Because I'm English teachers so... I have to improve my English"
- 4) "I got confidence in English- and I could tell my students I got Grade 1 when they asked me and I think not so many Japanese teachers have got first Grade in Japan"
- 5) "I wanted to pass Grade 1 so I am determined to study harder"
- 6) "No, nobody at staff is interested in my first grade"
- 7) "I had pressure to take Grade 1, because students have Grade 1"

Practicality

- 1) "I think EIKEN is helpful for motivating Japanese people to study but not a good proficiency test, but it's practical and not expensive, and easy to take and get textbooks to study so it's accessible."