The Author Responds: A Brief Clarification

Laura MacGregor
Sapporo International University

Nigel Henry presents some valid concerns about the test subjects, reliability measures, and validity findings in his critique of "The *Eiken* Test: An Investigation" (MacGregor, 1997). I will address them briefly here and attempt to clarify my purpose.

First, Henry correctly pointed out that the subjects for my study were older than the group the test was originally developed for. E*ikyo* stated that the pre-second level test was for high school level students; however, it also stated that it was "appropriate for a wide range of ages, from high school students to adults in Japan" (Nihon Eigo Kentei Kyokai, 1994, p. 8). In my paper, I reported that the majority of the pre-second level test-takers in June 1996 were high school students (227,666 or 75%). However, this number represents only 38.2% of the high school students who took the *Eiken* test during that test administration (.11% took the first level, .88% took the pre-first level, 16.8% took the second level, 37.1% took the third level, 6.3% took the fourth level, and .55% took the fifth level) (Nihon Eigo Kentei Kyokai, 1996, p. 11). Therefore, there is a disparity between *Eikyo's* ideal level of difficulty and the reality of the test-taking population. Following the above trend, I attempted to reflect the reality of the student population at my college, the subjects used in my study. The results of my student survey showed that, based on their test-taking experience, the pre-second level was the best choice (MacGregor, 1997, p. 28).

Second, Henry questioned my choice of reliability measures and implied that there should be CRT (criterion-referenced test) standards as well as the NRT (norm-referenced test) standards presented. This, however, would be impossible, because test reliability is based on test scores, and the *Eiken* test is scored as an NRT (i.e., it converts raw test scores to standardized scores) not as a CRT (in which test scores are interpreted as absolute).

Where test scores are concerned, NRTs and CRTs are completely different: NRTs aim to spread test scores over a wide continuum, and thus have a normal distribution and a high standard deviation. CRTs, on the other hand, aim to produce test results which have little variance,

that is to say, a low standard deviation. Since the *Eiken* test is administered to test-takers with wide ranging abilities and levels (high school, junior college, university, and post-university adults) and the relative scores follow a normal distribution pattern, it must be treated as an NRT, at least from a scoring point of view. To apply CRT reliability criteria would therefore be incorrect.

Henry's final area of inquiry questioned my validity study. He noted that *my discussion of poorly constructed items was* subjective, not empirical. It was my understanding that test validity *was* largely judgmental (Brown, 1996, pp. 231-239), and, therefore, I based my findings on a combination of the facts at hand: the aims of the test (as stated by *Eikyo*), current usage (Swan, 1995), and interpretations of the language as a native speaker of English.

Although the study has some shortcomings, I hope it will serve a larger purpose of alerting the people at *Eikyo* that there is some dissatisfaction with their public relations services and will encourage them to provide more information in the form of regular reports on the research and development of the *Eiken* tests.

References

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

MacGregor, L. (1997). The *Eiken* test: An investigation. *JALT Journal, 19*(1), 24-42.

Nihon Eigo Kentei Kyokai. (1994). *The STEP test?* Tokyo: Author.

Nihon Eigo Kentei Kyokai. (1996, September). *STEP news, no. 369.* Tokyo: Author.

Swan, M. (1995). *Practical English usage* (2nd ed.). Oxford: Oxford University Press.