

ABSTRACT

Rr. Regina Rani Widahadi. (2002). *Constructing a Criterion-Referenced Test to Assess Elementary School Students' Readiness in Following the English Instruction Presented in the Classroom: A Case Study on the Fifth Graders of SD Kanisius Kadirojo Yogyakarta*. Yogyakarta: Sanata Dharma University.

The presentation of the English instruction started from the elementary education level invited pros and cons among parents and scholars. The question of students' readiness in following the English instruction presented in the classroom was one among other considerations around the implementation of the educational policy.

In response to this issue, this study aimed at constructing a reliable and valid criterion-referenced test as an instrument to assess the elementary school students' readiness in following the English instruction presented in the classroom. In addition, it was intended to describe the results of the test if it is applied in the elementary school.

This was a case study on the fifth graders of *SD Kanisius Kadirojo Yogyakarta* in the third trimester of the academic year 2001/2002. The sample observed were 30 students. Pupils' *mastery* and *non-mastery*, the operational terms for the construct *readiness*, were specifically examined on the topic of *parts of the head*.

Test reliability was elaborated by (1) the *agreement coefficient* (ρ_0) obtained from the administration of the pre-and post-tests over a two-weeks-period which was 0.47 and (2) the *kappa coefficient* (κ) which was 0.12.

The degree of test validity was identified by the *expectancy table* which was comparing the post-test scores to students' *mastery* and *non-mastery* in the first formative test in the third trimester. The post-test scores were grouped into three groups of *above average* (≥ 8), *average* (7-7.9), and *below average* (≤ 6.9). The comparison showed 83.3% (25 pupils) in the above-average group (≥ 8) in the post-test demonstrated mastery in the first formative test in the third trimester. No student in the average (7-7.9) and below-average (≤ 6.9) groups demonstrated mastery since the cut-off score was at the point of 80% correct.

Test results were elaborated in two descriptive statistics of central tendency and dispersion. *Central tendency* described the mean (\bar{X}), the mode, the median and the midpoint. Those of the pre-test were 7, 10, 6.7, and 6. While those of the post-test were 8.9, 10, 9.5, and 6.6. *Dispersion* described the *range*, the *standard deviation*, and the *variance*. Those of the pre-test were 9, 2.3, and 5.3. While those of the post-test were 7.7, 1.8, and 3.3.

The analysis on students' ability to define 80% of the vocabularies of *parts of the head* in the pre- and post-tests was shown by the percentage-correct table. The table showed 14 students improved their achievement in the post-test. They could achieve the cut-off scores of 80% correct after following the instruction.

The item analysis revealing the effectiveness of items as the measure of *instructional effects* was as follows:

- Items 1,2,3,4,5,6,7,8,9,10,12,13,14,15 were effective ones, with fairly typical responses pattern. Some pupils responded correctly before instruction but a larger proportion did so after instruction
- Item 11 illustrated a defective one. Some pupils responded correctly before instruction but less proportion did so after instruction.

Concerning the aims of the study, and by basing on the findings, the conclusions were drawn. About test reliability, the decision consistency was about 47% and the test appeared to have low consistency. The percentage of classification agreement beyond chance was about 12%, therefore the test appeared to have low consistency. Test validity was expressed by the degree of relationship between the post-test and the first formative test in the third trimester which were 83.3%. The estimation showed that pupils who scored above average in the post-test would probably demonstrate mastery in the first formative test in the third trimester.

The computation of the pre- and post-tests scores revealing *central tendency* was elaborated in the *mean* (\bar{X}), the *median* and the *midpoint*. They indicated increases successively at the points of 1.9, 2.8 and 0.5. The *mode* of the pre- and post-tests was 10. The computation of the *range*, the *standard deviation*, and the *variance* as the indicators of the *dispersion* showed narrower value in the post-test. In two different ways the descriptive statistics showed improvement from the pre- to the post-tests. It implied the effectiveness of the instruction.

After all, it should be admitted that this study contained weakness in the identification of the degree of relationship between the criterion-referenced test constructed and the first formative test (the criterion-related validity). The value found was predictive since the two tests did not have exactly the same content validity. To avoid it in other studies, good communication and coordination between students doing a research study and teacher at school should be built.

To improve the reliability coefficients, the administration of the tests at seven in the morning may provide better time for the students. The tests in this study were administered at 9.30. Or, the test may be administered on a different population to obtain higher reliability coefficient. Therefore, a more reliable test will be fulfilled.

ABSTRAK

Rr. Regina Rani Widahadi. (2002). *Penyusunan Tes Criterion-Referenced sebagai Alat Ukur Kesiapan Murid-murid Sekolah Dasar dalam Mengikuti Pelajaran Bahasa Inggris di Kelas: Studi Kasus pada Murid-murid Kelas 5 SD Kanisius Kadirojo Yogyakarta*. Universitas Sanata Dharma Yogyakarta.

Pengajaran bahasa Inggris di sekolah dasar mengundang reaksi setuju dan tidak setuju di kalangan orang tua murid dan para ahli pendidikan. Keraguan tentang kesiapan murid dalam mengikuti pelajaran bahasa Inggris menjadi salah satu pertanyaan para orang tua dan ahli pendidikan.

Menanggapi hal itu, penelitian ini bertujuan menyusun tes *criterion-referenced* yang reliabel dan valid sebagai alat ukur kesiapan murid-murid sekolah dasar dalam mengikuti pelajaran bahasa Inggris di kelas. Penelitian ini juga bertujuan memberikan gambaran tentang hasil pemberian tes tersebut pada murid-murid sekolah dasar.

Penelitian ini merupakan studi kasus pada murid-murid kelas 5 SD Kanisius Kadirojo Yogyakarta pada catur wulan ketiga tahun ajaran 2001/2002. Subyek penelitian sejumlah 30 orang. Istilah *kesiapan* dalam studi ini mengacu pada penguasaan (*mastery*) atau ketidakpenguasaan (*non-mastery*) murid-murid terhadap materi yang diberikan. Materi yang dimaksud adalah materi *parts of the head* (bagian-bagian kepala).

Reliabilitas tes ditunjukkan oleh (1) *agreement coefficient* (ρ_0) yang diperoleh dari pemberian pre-tes dan pos-tes dengan jeda dua minggu yang menunjukkan koefisien 0.47 dan (2) *kappa coefficient* (κ) yang menunjukkan koefisien 0.12.

Tingkat kesahihan tes dibuktikan dengan *expectancy table* yang membandingkan nilai pos-tes dengan nilai ulangan harian pertama pada catur wulan ketiga. Kategori pembandingnya adalah penguasaan atau ketidakpenguasaan murid-murid terhadap materi yang telah diberikan. Dalam pembandingan ini nilai pos-tes dikelompokkan ke dalam kelompok-kelompok di atas rata-rata (≥ 8), rata-rata (7-7,9), dan di bawah rata-rata ($\leq 6,9$). Hasil pembandingan yang didapat adalah 83,3% (25 murid) yang berada pada kelompok di atas rata-rata (≥ 8) pada pos tes menunjukkan penguasaan materi pada ulangan harian pertama di catur wulan ketiga. Tidak ada murid dalam kelompok rata-rata (7-7,9) dan kelompok di bawah rata-rata ($\leq 6,9$) yang menunjukkan penguasaan materi pada ulangan harian pertama di catur wulan ketiga karena batas penguasaan materi adalah pada tingkat 80%.

Hasil pretes dan postes dinyatakan melalui teknik statistik sederhana (*descriptive statistics*) yang menggambarkan kecenderungan sebaran nilai (*central tendency*) pada kelompok yang diteliti dan sebaran nilai individu (*dispersion*) terhadap kelompok tersebut. Kecenderungan sebaran nilai (*central tendency*) meliputi mean (\bar{X}), mode, median dan midpoint. Mean (\bar{X}), mode, median dan midpoint pada pretes menunjuk pada angka-angka 7, 10, 6,7, dan 6. Mean (\bar{X}), mode, median dan midpoint pada postes menunjuk pada angka-angka 8,9, 10,

9,5, dan 6,6. Sedangkan sebaran nilai individu (*dispersion*) ditunjukkan oleh range, standard deviation dan variance. Range, standard deviation dan variance pada pretes menunjuk pada angka-angka 9, 2,3, dan 5,3. Range, standard deviation dan variance pada postes menunjuk pada angka-angka 7,7, 1,8, dan 3,3.

Analisa tentang penguasaan murid-murid dalam mendefinisikan 80% kosa kata *parts of the head* (bagian-bagian kepala) pada pretes dan postes ditunjukkan dalam tabel prosentase ketepatan jawaban. Tabel tersebut menunjukkan 14 murid mengalami peningkatan perolehan nilai pada postes. Mereka dapat mencapai batas penguasaan materi pada tingkat 80% setelah mengikuti pembahasan topik *parts of the head*.

Hasil analisa butir soal yang menunjukkan efektivitas butir soal sebagai pengukur efek-efek pengajaran (*instructional effects*) adalah sebagai berikut:

- Butir-butir soal nomor 1,2,3,4,5,6,7,8,9,10,12,13,14,15 merupakan butir-butir soal yang efektif, dengan pola yang biasa terjadi. Sejumlah murid menjawab dengan benar sebelum pengajaran dan kemudian semakin banyak murid yang menjawab dengan benar setelah pengajaran.
- Butir soal nomor 11 menggambarkan butir soal yang tidak sempurna dalam mengukur efek-efek pengajaran. Sejumlah murid menjawab dengan benar sebelum pengajaran dan kemudian hanya beberapa yang menjawab benar setelah pengajaran.

Berikut ini adalah kesimpulan yang didasarkan pada tujuan diselenggarakannya studi ini dan hasil-hasil yang telah diuraikan. Pengujian reliabilitas tes menyatakan besarnya konsistensi pengukuran tes sekitar 47%. Prosentase ini menunjukkan bahwa tes tersebut memiliki konsistensi pengukuran yang rendah. Prosentase kesesuaian pengelompokan di luar kebetulan sekitar 12%. Prosentase ini menunjukkan bahwa tes tersebut memiliki konsistensi pengukuran yang rendah. Validitas tes dinyatakan dengan tingkat validitas sejalan antara nilai postes dan nilai ulangan harian pertama pada catur wulan ketiga sekitar 83,3%. Artinya, dapat diperkirakan bahwa murid-murid yang berada pada kelompok di atas rata-rata pada postes akan menunjukkan penguasaan materi pada ulangan harian pertama di catur wulan ketiga.

Penghitungan nilai-nilai pretes dan postes menunjukkan peningkatan pada kecenderungan sebaran nilai (*central tendency*) yang dinyatakan dalam mean (\bar{X}), median dan midpoint berturut-turut sebesar 1,9, 2,8 dan 0,5. Mode pada pretes dan postes adalah 10. Penghitungan nilai-nilai pretes dan postes menunjukkan penyempitan pada sebaran nilai individu (*dispersion*) yang dinyatakan dalam range, standard deviation dan variance. Melalui dua cara yang berbeda, teknik statistik sederhana (*descriptive statistics*) tersebut menunjukkan peningkatan dari pretes ke postes. Temuan ini menunjukkan keefektivan pengajaran.

Akhirnya perlu dinyatakan tentang adanya kelemahan dalam studi ini. Nilai tingkat keterhubungan antara nilai postes dan penguasaan materi pada ulangan harian pertama di catur wulan ketiga yang ditunjukkan dalam studi ini lebih merupakan nilai perkiraan. Ini disebabkan kedua tes yang dibandingkan tidak memiliki kesamaan validitas isi. Untuk menghindari kelemahan ini pada studi serupa, perlu dibangun komunikasi dan koordinasi yang baik antara

mahasiswa yang melaksanakan studi penelitian dan guru yang mengajar di sekolah.

Untuk meningkatkan koefisien reliabilitas tes perlu dipertimbangkan waktu pemberian tes yang tepat. Pemberian tes pada pukul tujuh mungkin merupakan waktu pemberian tes yang lebih tepat bagi murid kelas 5 sekolah dasar mengingat tes pada studi ini diberikan pada pukul 9.30. Sedangkan pemberian tes kepada populasi lain memungkinkan pemerolehan reliabilitas yang lebih tinggi, yang menunjukkan tes yang lebih reliabel.

