Assessing the Validity and Reliability of a Paper-and-Pencil Piagetian Test

Amy Dugan

Department of Counseling Education and School Psychology, Wichita State University, Wichita, KS, 67260, USA

Introduction

In 2004 37.2% of Wichita Public Schools tenth grade students scored proficient or above on the state math assessment; seventh grade math scores were not much greater: 48.2% (Funk, 2005). Reading and writing scores for middle and high school students ranged from 53% to 65%. The way a person learns and processes information is dependent upon their cognitive ability (Crain, 2005); people cannot learn information that is presented in a manner that is too complex for their level of cognitive understanding. If teachers knew in what stage of cognitive development their students were operating, they could plan developmentally appropriately lessons.

Piaget hypothesized that children learn in qualitatively different ways based upon their level of development. He placed people in one of four stages: sensorimotor, preoperational, concrete, and formal operations (Muuss, 1996). In order to determine what stage of cognitive development a person was operating in, Piaget developed a series of tasks which he used to assess children’s levels of cognitive abilities. The clinical method developed by Piaget is administered in a one-on-one setting; it is not a method which is feasible for use by a classroom teacher. Thus, there is a need for an objectively scored paper-pencil test that classifies cognitive abilities which could be administered by classroom teachers to groups of students. Bakken (1995) created a paper-pencil test based on Piaget’s tasks that can be used by classroom teachers to determine what level(s) of cognition their students are functioning in. Unfortunately, problems with the scoring technique limited the validity and reliability of the instrument; therefore, a revised scoring system was developed that simplified and broadened the scoring of the test.

Hypothesis

There are three hypotheses in this study. First, the criterion-related validity of the instrument will be increased due to restructuring of the scoring method. Second, the test-retest reliability of the instrument will increase due to restructuring of the scoring method. Third, the difference between the old and new scoring methods will be significantly statistically different.

Experiment, Results, Discussion, and Significance

Participants

This study is using data that was collected on participants during a previous study. Third, fifth, seventh, tenth, and twelfth grade students participated. Some attended parochial school while others attended public.

Instrument

There are 21 multiple-choice items on the objectively scored Piagetian based paper-pencil test that was designed to determine what stage of development children are operating in (Bakken, Thompson, Johnson, Cladywer, 2001). The test can be used to classify students as preoperational, concrete operational, or formal operational; this test also assesses which substages children are operating in (see Appendix B).

Two criterion-related validity studies of the instrument have been conducted (Skar, Bakken, Thompson, Johnson, 1995). A random sample of 40 third grade students from four elementary classrooms participated in the first study. Each child was given the concrete operational Piaget tasks using a clinical interview technique; then the students were administered the written test for those items that measured concrete operations. The paper-and-pencil test indicated a statistically significant correlation for concrete operations: sub 1 (r = .75, p < .01), sub 2 (r = .69, p <
.01), and sub 3 (r = .69, p < .01). For the second validation study, a random sample of 40 fifth grade children from four classrooms were assessed. Concrete operational substages indicated a significant correlation (r = .69, p < .01), and both formal operation substages also yielded a significant correlation (r = .54, p < .01 for the first substage; and r = .45, p < .05 for the second substage). Thus, the multiple-choice test of Piaget tasks suggests a valid technique of identifying individuals’ cognitive thinking, and offers a means of administration to groups at one sitting; however, restructuring of the scoring method may increase the validity of the instrument.

In order to establish test-retest reliability two studies were done which measured students’ responses to the Piaget paper-pencil test at a 3 week interval (Skar, et al., 1995). The first study included 80 seventh grade students. Statistically significant correlations were found for concrete operations (sub 1, r = .24, p < .05; sub 2, r = .37, p < .01; sub 3, r = .42, p < .01), and for formal operations (sub1, r = .44, p < .01; sub 2, r = .54, p < .01). The second study assessed a total of 56 students from ninth and eleventh grades. Data indicated a significant correlation for concrete operations (sub 1, r = .67, p < .01; sub 2, r = .63, p < .01; sub 3, r = .69, p < .01) for formal operations (r = .70, p < .01 for the first substage and r = .61, p < .01 for the second substage). Data suggests that the Piaget paper-pencil test is a reliable means of identifying students’ levels of cognitive thinking; however, restructuring of the scoring method may strengthen the reliability of the instrument.

Procedure

Bird (2005) conducted a study of seventh-grade children, using Bakken’s (1995) Piagetian paper-and-pencil test. However, in scoring the protocols she found three items that were missed by over 85% of the students in the beginning formal stages, but students would then get later-stage formal operation items correct that were successively more difficult. Consequently, items 14, 15, and 16 were eliminated from the new scoring procedure because of extreme item difficulty and stage-like order of development that was missing.

Data Analysis

In order to establish test-retest reliability a correlational analysis of the new scoring method will be used to determine the reliability between the pre-test and the post-test scores. In order to determine criterion-related validity, we will use a correlational analysis between paper-pencil tests and clinical interviews. For the third hypothesis, a t-test after Hotelling will be used to analyze whether the validity and reliability coefficients of the revised scoring method are statistically significantly different than the original scoring method.

Conclusions

The revised scoring method found stronger relationships between test and retest scores than the original scoring method did. Validity correlations were also stronger when the revised scoring method was used. The t-test indicated that the revised scoring method was statistically significantly more reliable for each substage analyzed; however, it was only statistically significantly more valid for two of the three substages.

References


