# Beery-Buktenica Visual Motor Integration Test

## Validity Evidence Examination

*Rosa Román-Oyola*
*Víctor E. Bonilla Rodríguez*

RESUMEN

Este estudio exploratorio examinó la validez de los resultados obtenidos mediante la administración de la Beery VMI a una muestra de niños puertorriqueños. El proceso incluyó la traducción de las instrucciones de la prueba, entrevistas a cuatro profesionales que utilizan la prueba y la administración de la prueba a 50 estudiantes de Kindergarten de dos escuelas de San Juan, Puerto Rico. Los datos cuantitativos y cualitativos se recopilaron a base de los Estándares para las Pruebas Educativas y Psicológicas (1999): contenido, proceso de respuesta, estructura interna, relación con otras variables y consecuencias de la prueba. La información respecto al contenido, el proceso de respuesta y las consecuencias de la prueba apoyaron la validez de los resultados. Sin embargo, se deben realizar otros estudios para establecer una conclusión definitiva.

**Descriptores:** Beery VMI, validez, integración visual motora

ABSTRACT

This exploratory study examined the validity of the results obtained from the administration of the Beery VMI to a sample of Puerto Rican children. The process included translation of test instructions, interviews with four professionals that use the test in their clinical practice, and the administration of the test to 50 Kindergarten students from two schools in San Juan, Puerto Rico. Quantitative and qualitative data was collected as established by the Standards for Educational and Psychological Testing (1999): test content, response process, internal structure, relation to other variables, and test consequences. Information regarding test content, response process, and test consequences supported the validity of the results. However, more research is needed to reach a definitive conclusion.

**Keywords:** Beery VMI, validity, visual motor integration

## Introduction

Handwriting development is one of the most emphasized areas during preschool education. According to McHale and Cermark (1992), children from public schools spend around 30 to 60% of the day working on fine motor activities. It is estimated that 85% of the time dedicated to those tasks involves activities using paper and pencil. Authors from different disciplines have researched which skills a child needs in order to attain a legible and good handwriting. Findings suggest a relationship between visual motor integration skills and the legibility and quality of handwriting (Tseng & Murray, 1994; Cornhill & Case-Smith, 1996; Case-Smith, 2002; Daly, Kelley, & Krauss, 2003; Marr, Windsor, & Cermark, 2001; Weintraub & Graham, 2000). Hsu (1997) argued that one of the instruments most used internationally to evaluate children with visual-perceptual-motor difficulties is the Developmental Test of Visual Motor Integration (known as the Beery VMI on its 2004, fifth edition). The test consists of a developmental sequence of geometric figures to be imitated or copied with paper and pencil. It is accompanied by two supplemental subtests which evaluate the visual perception and motor coordination skills.

The literature reviewed supports Hsu's claim about the international use of the Beery VMI by psychologists, learning disabilities specialists, teachers, and occupational therapists. Beery and Beery (2004) state that the Beery VMI is virtually culture-free. Indeed, they say that the norms of the fourth edition of their test are commonly considered to be international norms. According to them, "It is hypothesized that the Beery VMI and its supplemental tests are not biased in regards to gender, ethnicity, socioeconomic status, and residence" (Beery & Beery, 2004, p.115). However, they advise that "Ideally, every test should be locally normalized on every population for which it will be used" (Beery & Beery, 2004, p. 89).

Richardson (2001, p. 240) explains that in recent years there has been a lot of criticism related to the "validity of tests developed primarily on a white, middle-class population when they are used with children from different cultural backgrounds". Authors from the research methods (McMillan, 2004) and educational measurement fields (Allen & Yen, 2002) have informed that the norms and results obtained from one study apply only to the particular population involved in the study. Therefore, it is considered necessary to examine if the scoring norms presented on the Beery VMI, fifth edition manual, are applicable to children outside the United States with different cul-

tures and ethnic backgrounds. In this case, the Beery VMI was administered to a sample of Puerto Rican children to study the validity of its results using as a guide Messick's (1989) definition of validity and the Standards for Educational and Psychological Testing published in 1999 by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME).

## Purpose of the study and its relationship to the concept of validity

The purpose of this study was to examine the validity of the results obtained with a sample of Puerto Rican Kindergarten children on the Beery VMI when their performance is interpreted using the norms presented in the test's fifth edition manual.

Messick (1989) presents a broader and unified vision of validity by establishing that there is only one type of validity, which is based on different sources of evidence. According to this author, validity refers to the degree to which empirical and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores. For Messick, construct validity serves as a central core since it entails the other aspects of validity (e.g., content validity, criterion validity).

The definition of validity presented in the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999) is consistent with Messick ideas. According to the Standards, there are five sources of evidence that should be used to evaluate the proposed test scores interpretations considering the purpose for which the test was developed: 1) Evidence based on the content of the test refers to the relationship between test content and the construct it is supposed to measure; 2) The analysis of evidence based on the response process provides evidence of the correspondence between the construct and the nature of the performance or response offered by an individual; 3) Evidence based on the internal structure of the test refers to the degree at which the relationship among the items and the test's components correspond to the construct on which the interpretations of test scores are based; 4) Evidence based on the relationship with other variables involves the analysis of the relationship between test scores and other variables. These variables can be some criterion that will be predicted based on the test scores. They can also refer to some other tests with constructs related to the constructs of the test being evaluated or some other vari-

ables that could be relevant depending on the theory on which the test is based; and 5) Evidence based on test consequences emphasizes the importance that the items of the test really measure the proposed construct (and not other concepts or characteristics of the person) since there are decisions that will be made based on the results of the test.

## Psychometric characteristics of the Beery VMI

Beery and Beery (2004) mention studies that provide information on the psychometric characteristics of the Beery VMI test. In regards to the internal consistency of tests' results, they establish that "Because the Beery VMI results have remained quite consistent for many years, early studies are considered applicable to the present edition" (Beery & Beery, 2004, p. 99). For the group of children between four and six years, the Cronbach Alpha values range from .81 to .84 on the Beery VMI Main Test, .86 to .87 on the Visual Perception Subtest, and .84 to .88 on the Motor Coordination Subtest.

To support the validity of the Beery VMI results, its authors refer to different studies that date back to the first edition of the test in 1967. To evaluate the construct validity, they mention some hypotheses that have been supported by research. One of these hypotheses is the correlation, at least moderate, between the Beery VMI subtests and its Main Test of Visual Motor Integration. All the correlations calculated using participants' scores in the normalization process were significant at the .05 level of significance (Beery & Beery, 2004). The average correlation values obtained for the group of children with an average age of five were: .45 between the Main Test and the Visual Perception Subtest; .46 between the Main Test and the Motor Coordination Subtest; and .37 between the Visual Perception and Motor Coordination Subtests (Beery & Beery, 2004).

Kulp and Sortor (2003) studied the clinical value of the Beery VMI subtests to examine how the results of the different test's components correlate with each other. Among other findings, they concluded the following: (a) there is a significant correlation among the results of the different components of the test; (b) the lineal regression analysis indicated that the subtests' results explained 36% of the variation on the Main Test (VMI) results.

On the other hand, Beery and Beery (2004) allude to different studies which findings support that there is no statistically significant difference in terms of gender, ethnicity, socioeconomic status, and residency. For example, in regards to ethnicity, they mention a study from

Italy where no significant difference was found between children's performance and the one expected according to the normalization of the test (Beery & Preda, 2000, in Beery & Beery, 2004). In addition, according to Beery and Beery (2004), studies performed in China during 1972 and 1991 evidenced that, at early ages, Chinese children performed somewhat better than US children. The opposite happened with children from Greece (Georgas, 1971, in Beery & Beery, 2004) and Norway (Sovik, 1975, in Beery & Beery, 2004), who performed slightly below the average for US children. According to Beery and Beery (2004), only one study found statistically significant differences between the performance of African-American and Caucasian children (Nye, 1977, in Beery & Beery, 2004). However, a later study evidenced the opposite (Schooler, 1979, in Beery & Beery, 2004).

## Other demographics to be considered

The culture and ethnicity of the people from the normative sample of any standardized test are aspects that influence its validity when applying its results to populations from other countries. Hsu (1997) compared the performance on the Beery VMI of a sample of Taiwanese children with the expected results established with the normative data of the third edition of the test. According to Hsu's (1997) results, the norms of the third edition of the Beery VMI may be applied to Taiwanese children at first, fourth, and sixth grade. Earlier studies (Liu, 1972 and Mao, 1995, in Hsu, 1997), however, evidenced that, apparently, at early ages, Chinese children perform better when their results were compared to those of the normative sample of American children.

Other studies have provided interesting results regarding the visual motor integration skills and the socioeconomic status of children. Frey and Pinelli (1991) compared the performance on the Beery VMI and other visual motor integration tests of two groups of Brazilian children from different socioeconomic status. They found that kids from a low socioeconomic status obtained lower scores when compared with kids from a high socioeconomic background. Studies performed in other countries have found similar results pointing out that samples of children from low socioeconomic status at South Africa (Verdonck & Henneberg, 1997) and the South-East of the US (Baowman & Wallace, 1990) have performed poorer when compared with children form high socioeconomic status.

Marriott (2000) studied the relationship between visual motor integration skills and some other demographic variables such as age, gender, ethnicity, income, parents' educational level and previous school experience. Marriott (2000) did not use the Beery VMI in her research. She used one of the subtests included as part of the Kindergarten Diagnostic Instrument (KDI-II). Her findings indicated that ethnicity and income did not significantly influence children performance on the visual motor integration test. On the other hand, variables such as age, gender, and previous school experience, seemed to be influential factors. The performance of older children, girls, children with previous school experience, and children with parents of higher educational levels was better.

## Methodology

This study can be labeled as exploratory due to the small sample size used ($n = 50$). It consisted of a methodological research design in which different strategies were used to examine the reliability and validity of an instrument and determine its applicability scope (Portney & Watkins, 2000, p. 352). Table 1 presents the sources of information, their alignment with the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999), and the research questions.

## Population and sample

The population for this study consisted of Kindergarten children, ages 5 to 6 years from three public elementary schools from the San Juan metropolitan area. Demographic information was studied in different schools in order to select two that had different student body composition, thus maximizing the variation of students' characteristics in the sample. These schools came from different school districts, San Juan II and San Juan III. The selection of the groups of children fell mainly on the school principals and the Kindergarten teachers. One group was selected from each school for a total of 50 children. Schools were visited to present the study to the teachers, meet the children, and leave the parents' consent forms. Parents who consented to their child's participation in the study were required to fill out a child's demographic information sheet that included: gender, birth date, preschool experience, and parents' educational level. The test was administered individually, in a space separated from the classroom to avoid interruptions and distractions.

Table 1
Sources of evidence based on the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999) and their corresponding research questions

| *Source of evidence* | Data collection / analysis | Related research questions |
|---|---|---|
| *Test content* | • Interviews to four professionals. | 1. What is the opinion of four professionals (two psychologists and two occupational therapists) who regularly administer the Beery VMI in regards to the pertinence of its content according to the purposes for which they use it? |
| *Response process* | • Cognitive interviews performed to six Kindergarten students from a school in San Juan. | 2. Which characteristics of the children cognitive process are observed while they complete the Beery VMI?<br>3. Which modifications (according to what the test's manual allows) could be done during the administration of the Beery VMI to the Kindergarteners to make the test more appropriate for them? |
| *Test's internal structure* | • Correlation coefficients hypothesis test. | 4. How do the Beery VMI Cronbach's Alpha reliability coefficients calculated with the sample scores compare to those presented in the test's manual for the corresponding age group? |
| | • Correlation coefficients hypothesis test. | 5. How do the Pearson correlation coefficients calculated with the sample scores between the Beery VMI and its subtests compare to the Pearson correlation coefficients presented in the test's manual for the corresponding age group? |
| | • Multiple regression analysis. | 6. What percentage of the children's performance variance in the Main Test could be explained by their scores in the Visual Perception and Motor Coordination Subtests? |
| *Relationship with other variables* | • T tests and One way Analysis of variance using demographic data as independent variables. | 7. How does children performance varies in regards to personal variables such as: gender, socioeconomic status, previous preschool experience and parents' educational level? |
| *Test consequences* | • Z tests. | 8. How do the sample scores compare with the norms established on the Beery VMI fifth edition manual? |
| | • Interviews to four professionals. | 9. What is the opinion of four professionals (two psychologists and two occupational therapists) who regularly administer the Beery VMI in regards to the validity of the use of its scores' interpretations when the test is administered to Puerto Rican children? |

In addition to the sample of children, four professionals (two psychologists [P] and two occupational therapists [OT]) were selected using a purposive sampling strategy. These professionals participated from a 45 minute interview to gather information regarding content and test consequences based evidence. Interviews were recorded, transcribed, and sent to each person for verification purposes.

## Data collection and analysis

Data collection included both qualitative and quantitative data. All quantitative analyzes used an alpha of .05. The following sections present the processes used to analyze the data related to each source of evidence collected in this study.

*Test translation.* Test's instructions were translated to Spanish using a conceptual with backward translation process to capture the implicit associations or written text meaning in the instrument's original language (Braverman & Slater, 1996). Instructions were the only thing that needed to be translated, for test content is based on drawings and forms that the child needs to copy (in the Main Test), identify as equals (in the Subtest of Visual Perception) and trace within the lines (in the Subtest of Motor Coordination).

Six children, 3 boys and 3 girls, were selected from one school to perform cognitive interviews allowing the verification of the instrument's instructions translation, the instrument's administration, and providing evidence of the response process.

*Content based evidence.* This evidence stem from the semi-structured interviews performed to the professionals regarding their perception of test content pertinence related to the use they give to it. The categories that emerged from the process were: 1) purpose of the use, 2) content pertinence, 3) reasons / advantages of its use, 4) concerns regarding the test, 5) ways in which test results are interpreted, 6) clinical observations made when using the test, 7) other criteria used to interpret children's performance, 8) use of test results / decisions based on test results, and 9) comments regarding the use and interpretation of the test's normative data with Puerto Rican children.

*Response process based evidence.* Data was collected through cognitive interviews performed to six Kindergarten children. According to Fowler (2002), these interviews are performed to examine how participants understand and respond the questions contained in an instrument. Therefore, this process inquired if children understood the instructions and what they thought while performing similar tasks. This data was analyzed to test administration procedures.

*Test's internal structure.* Data collected through the Beery VMI administration was entered into a Statistical Package for the Social Sciences (SPSS) database. Item responses were coded as 1 (correct) or 0 (incorrect). Also, the zero score was used for items that were not asked to students, due to them reaching the test discontinuation criterion (three sequential incorrect responses).

Different procedures were performed to analyze data related to the test's internal structure: reliability coefficients comparison, correlation coefficients comparison, and a regression analysis. The researchers tried to perform a confirmatory factor analysis on the items of the Beery VMI and its subtests to examine how the items of the three test components (Main Test and subtests) grouped together in relation to the underlying test constructs (visual-motor integration, visual perception, and motor coordination). However, the analysis could not be performed because there was not enough variability (Stevens, 2002). This problem could be due to a combination of two factors: the way in which the test is scored and the sample size.

The first analysis performed was a Cronbach's Alpha reliability coefficients comparison between those obtained from the study and those reported in the test manual for the corresponding age group. The researchers decided to calculate Cronbach's Alpha to enable the comparison with the information presented by Beery and Beery (2004). The test selected provides a way to determine if there are any differences between the strength of two correlation coefficients. It involves the transformation of the coefficients using a Fisher's Z transformation formula. The test was performed using the online service provided by the Institute of Phonetic Sciences of Amsterdam (http://www.fon.hum.uva.nl/Service/Statistics/Two_Correlations.html).

The second analysis involved the comparison of the correlation coefficients reported by Beery and Beery in the test manual for the Beery VMI and its subtests for the corresponding age group and those obtained from the study. The Pearson correlation coefficient was calculated for each pair formed by the Main Test and the supplemental tests. Pearson was used because the scores being correlated were the total scores obtained by each student in each test pair. The online service provided by the Institute of Phonetic Sciences of Amsterdam was used to compare the strength of the correlation coefficients obtained in the study with those reported by Beery and Beery (2004).

The third analysis involved a multiple regression analysis to calculate what percentage of variability in children's performance in the Main Test was explained by the results obtained in the supplemental tests of visual perception and motor coordination. This analysis was conceived following Beery and Beery statement that the skills measured by the Main Test and those measured by the subtests are related, because each subtest measures a part of what is measured in the Main Test (2004, p. 106). The multiple regression analysis used the Main Test

as the criterion and the supplemental tests as predictors (Visual perception and Motor coordination). The analysis was performed using SPSS. A backward inclusion method was selected, because this method allows including all variables in the first model. The amount of shared variance and the predictive power of each variable (Beta weights) were evaluated to determine the resulting equation.

*Evidence based on the relationship with other variables.* The Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999) state the importance of gathering evidence on how tests results behave in relation to other variables. Variables considered in this part of the study were: gender, socioeconomic status, parents' educational level, and previous school experience.

Independent Samples t Tests were performed to examine whether there were any differences on test results (the dependent variable) according to gender, previous school experience, and socioeconomic status (the independent variables). The alternative hypotheses were non directional, for the literature reviewed did not provide conclusive information regarding a possible direction for the results. All tests used an alpha of .05 to determine significance. Cohen's *d* coefficient was used to determine effect sizes. Effect sizes were reported using Hedges correction for bias due to small sample sizes (Grissom & Kim, 2005). According to Cohen (1988), effect sizes can be classified as small if around .2, medium if around .5, and large if .8 or more. The importance of calculating effect sizes lies in that "a test of statistical significance provides the quantified strength of evidence (attained *p* level) that a null hypothesis is wrong, while an effect size (ES) measures the degree to which such null hypothesis is wrong" (Grissom & Kim, 2005, p. 4).

A One-way Analysis of Variance (ANOVA) was performed to test for differences among groups formed with the variable parents' educational level. The levels for the variable were selected according to those presented by Beery and Beery (2004) as part of the normative sample demographic characteristics. The alpha level selected to determine significance was .05. Omega squared was used to determine the effect size. Effect sizes computed with omega square can be classified as small if around .01, medium if around .06, and large if .14 or more (Cohen, 1988).

*Evidence based on test consequences.* Two sources of information were used to gather evidence on the consequences of test use. The Berry VMI is supposed to be used as part of a battery of tools that professionals chose to administer to determine children's need of treat-

ment or of a follow up program for home or school to facilitate their performance in visual motor integration activities. Therefore, those decisions (need of treatment or follow up program) are considered as the consequences of this test (i.e., they are what comes for the child after the administration of the Beery VMI). Currently, performance of Puerto Rican children on the Beery VMI is interpreted based on the manual's norms (established with children residing in the US).

Since this study is looking for evidences of validity, it was considered pertinent to compare our sample's performance in the Main Test and its subtests to the norms presented in the manual. This allowed the evaluation of test's consequences by determining if the means established as norms in the test's manual (on which professionals based the decisions they made) are comparable to the means calculated with the results of the preschoolers sampled in this study. With that purpose, *z* tests were performed. According to the manual, the population mean is 100 in all three tests, while the standard deviation is 15. A non directional alternative hypothesis was established for the reviewed literature did not provide any indication of a possible direction for the test's results.

The second source of information regarding test's consequences was gathered from four semi-structured interviews performed to professionals that use the Beery VMI in their clinical practice (two psychologists and two occupational therapists). Interviews collected information on test use and interpretation, decisions made based on test results, and other criteria used to determine the level of visual-motor integration skills of the children, among other aspects.

## Results

The average age of the children who participated in this study was 69 months (5 years and 9 months). There were 25 boys and 25 girls. Eighty-eight percent ($f = 44$) had previous school experience. Table 2 shows that the amount of boys and girls in both schools was similar. The majority of the children in both schools (92% in the school with high socioeconomic status and 84% in the one with low socioeconomic status) had school experience prior to Kindergarten.

Most parents' ($f = 42$) reported having an educational level of one through 3 years of college education. However, in the school with high socioeconomic status there were more parents with "1-3 years of college" (56% of the fathers and 48% of the mothers), and in the school with low socioeconomic status the largest amount was found in the category of "High School" (52% of the fathers and 52% of the mothers).

Table 2
*Sample description in terms of children's school of origin (n = 25 for both schools)*

| | School of origin | | | |
|---|---|---|---|---|
| | High Socioeconomic Status | | Low Socioeconomic Status | |
| | *P* | *f* | *P* | *f* |
| Gender | | | | |
| ▪ Feminine | 52 | 13 | 48 | 12 |
| ▪ Masculine | 48 | 12 | 52 | 13 |
| School experience previous to Kindergarten | | | | |
| ▪ Yes | 92 | 23 | 84 | 21 |
| ▪ No | 8 | 2 | 16 | 4 |
| Father's educational level | | | | |
| ▪ < High School | 4 | 1 | 16 | 4 |
| ▪ High School | 20 | 5 | 52 | 13 |
| ▪ College 1-3 years | 56 | 14 | 32 | 8 |
| ▪ College 4+ Years | 20 | 5 | 0 | 0 |
| Mother's educational level | | | | |
| ▪ < High School | 4 | 1 | 12 | 3 |
| ▪ High School | 12 | 3 | 52 | 13 |
| ▪ College 1-3 years | 48 | 12 | 32 | 8 |
| ▪ College 4+ Years | 36 | 9 | 4 | 1 |

*Notes*. Educational levels are the same presented in the Beery VMI 5th ed. manual (p. 98); *P* = Percentage; *f* = frequency.

*Content based evidence.* The results obtained from the four semi-structured interviews indicated that all professionals used the Beery VMI to evaluate visual-motor, visual perception, and fine motor coordination skills. The psychologists emphasized the academic aspect as one that they relate to the child's performance in the Beery VMI. For example, P1 stated the impact of the skills evaluated by the Beery VMI in "reading, writing, and mathematics, in summary, in their academic achievement". Both psychologists agreed on the usefulness of the test to detect difficulties in the areas that the test evaluates.

In regards to the pertinence of the content, professionals noted different test characteristics. Two of them talked about the way in which the test requires children to work in order to measure their visual-motor integration:

> OT2--- I think that the content of the test is appropriate because it is specific enough in terms of what the child needs to do when copying the drawings.

> P2--- …I would say that, in some respect, the content of the test is adequate, because if the child is looking at a pattern, he or she needs to process and reproduce it. We are using everything: the mental map, memory, and the ability to reproduce manually…

On the other hand, one OT emphasized the pertinence of the way in which items are organized, indicating that it helps children to visually organize the task at hand. This therapist also stated that the content that has been added to the most recent version of the Berry VMI is appropriate.

> OT1--- I like (for example), in the visual perception part that the child needs to point at different parts of the body… I like this because it does not penalize a lot. That is, it helps me to see if the child has learned some concepts even when there are others that he or she still has not acquired... It is the same with the part dealing with motor skills (because it gives credit to the child for easier things) in terms of how he or she handles the pencil, how he or she traces… and the other part (the Main Test) is easy to work with because it is very basic. It is not complex.

P1 stated that it is an instrument that the child can work: "The designs are not difficult at the beginning. Obviously, things change as the test progresses. But, my experience has been that it is received well".

Apparently, the fact that the test is manageable is one of the reasons for these professionals to use the Beery VMI. They agreed in that the advantages of the Beery VMI are: easy to administer, fast, and provides the opportunity to make clinical observations while the child performs the tracing task. Other advantages include that the test is well known, they feel comfortable administrating it, and that it is relatively easy to explain the results to the parents.

> OT2--- …It is a well known test. Most professionals know what is the VMI, especially the psychologists who administer it… and the way to explain the results to the parents is easy as well…

> P2--- It is faster, more structured, and the child can work by him or herself… That gives me more time to observe more closely other aspects of his or her performance…

However, in spite of the advantages, they also expressed worries regarding three aspects: the validity of test results (specifically for preschool children), the need to include a guide for clinical observations as part of the test, and the lack of information regarding the congru-

ency of results between the Beery VMI and other tests that measure similar constructs.

> OT1--- …Between the ages of three and five, the child tends to do well… But as the child gets older, parents ask me why he or she obtained lower scores… The reason is that these children have mastered these skills. It is what they do all the time. But when they go to first grade, the demands increase. They have to copy from the board, take dictation, etc. Therefore, I sometimes think that the test does not take those things into consideration.

> OT2--- The only thing that worries me about the test is that it should have an appendix, something dealing more with clinical observations like how the child performed, how was his or her conduct, where he or she starts to trace, where does he or she end…

> P2--- …I have had difficulty or have noted, and other colleagues with which I had spoken have said the same thing, is that sometimes, when one is correcting the test there is no congruency with the test of intelligence (particularly in those sections in which the child needs to, for example, copy geometrical figures)…

*Evidence based on the response process.* Cognitive interviews focused in gaining information regarding the understanding of instructions and on what children thought while looking at and working with each drawing. Most children's responses reflect that they made reference to the name of the figure they were observing. However, some associated the shapes with one of the letters of their names, while others refer to objects familiar to them, for example: an add sign or God's cross (item 10 -cross), a blackboard or a truck (item 12 - square), a slide (items 11 and 13 - diagonal lines), a mountain (item 15 - triangle), the number four without a line (item 16 - square with the top opened and a circle), the whiskers of a kitten (item 17 - asterisk), a street sign that indicated which way to go (item 18 - cross with arrows), and three girls playing with "hula hoops" when viewed from above (item 19 - three overlapping circles).

Cognitive interviews flowed smoothly. Children expressed understanding of the instructions and were able to paraphrase and explain them. Also, they seemed to enjoy the process and try to do their best. Therefore, the instructions and the administration process were considered to be adequate.

*Evidence based on the test's internal structure.* Cronbach's Alpha internal consistency coefficients were examined individually and then

compared to those presented in the test's manual (see Table 3). All reliability coefficients presented by Beery and Beery are .8 or more. However, the only coefficient equal to .8 among those calculated based on the sample was found for the Visual Perception Subtest. According to Crano and Brewer (1973), good reliability coefficients are .8 or larger. Based on this criterion, the only coefficient that was adequate was the one obtained for the Visual Perception Subtest.

Table 3
*Cronbach's Alpha reliability coefficients calculated with the sample scores and those presented in the test's manual*

| | **Sample** | **Beery and Beery (2004)** |
|---|---|---|
| *Main Test* | .61 | .84 |
| *Visual Perception Subtest* | .80 | .87 |
| *Motor Coordination Subtest* | .56 | .88 |

A plausible explanation for obtaining low reliability coefficients is found when examining the number of items in each test versus the number of cases. First, SPSS eliminated automatically some items which variability was zero from the analysis. Literature states as a general rule that the greater the amount of items, the larger the reliability coefficient. Also, the larger the amount of items, the greater the sample size needed to perform the analysis and obtain good reliability coefficients.

Once SPSS eliminated items from the analysis, the Main Test was left with 10 out of 21 items, the Visual Perception Subtest was left with 22 out of 30, and the Motor Coordination Subtest was left with 15 out of 30. In regards to sample size, Beery and Beery indicated that the sample size in their study was of 192 children, while this study had only 50 children.

After stating whether the coefficients obtained were appropriate according to the literature, the researchers proceed to test whether there were significant differences between the coefficients obtained and those reported by Beery and Beery (2004). Significant differences were found for the coefficients of the Main Test and the Motor Coordination Subtest ($p$=.00167 and $p$= 5.17e-06, respectively). No significant difference was found for Visual Perception ($p$ =.15).

The second analysis involved the calculation of Pearson correlation coefficients among each of the test's components (Main Test and

subtests). Hypothesis testing was performed on these correlation coefficients comparing them to those presented in the test manual (see Table 4). Following the guidelines established by Hinkle, Wiersma, and Jurs (2003), all coefficients (both obtained in this research and those presented by Beery and Beery) are low and positive. This finding supports Beery and Beery (2004) statement in that the Main Test measures components that are related (visual perception and motor coordination), therefore a positive correlation among the tests. However, it does not measure exactly the same components, for one of the subtests measures visual perception skills while the other measures motor coordination skills.

Table 4
*Pearson correlation coefficients calculated with the sample scores and those presented in the test's manual*

| | **Sample** | **Beery and Beery (2004)** | ***p values*** |
|---|---|---|---|
| *Main Test* vs. *Visual Perception Subtest* | .41 | .45 | .763 |
| *Main Test* vs. *Motor Coordination Subtest* | .37 | .46 | .504 |
| *Visual Perception Subtest* vs. *Motor Coordination Subtest* | .43 | .37 | .661 |

*Notes*. Sample $n = 50$; Beery and Beery $n = 192$.

Non significant differences were found between the correlation coefficients obtained in this study and those obtained by Beery and Beery. Therefore, it can be stated that the coefficients obtained in this study are similar to those obtained by Beery and Beery (2004).

The third analysis used a multiple regression to provide more evidence on the internal structure of the tests. This analysis aimed to identify what percentage of variability in the Main Test is accounted for by the subtests. According to Tabachnick and Fidell (2007), in order to perform this analysis one has to meet two very important assumptions: the absence of singularity (the Main Test cannot be a combination of the subtests) and the absence of collinearity. Since each test includes different items, singularity is not a concern. Collinearity was checked using different indexes none of which evidenced its presence. Other assumptions regarding the multiple regression analysis were met as well: multivariate normality, linearity, and homocedasticity.

The regression analysis results indicated that the model explained 21.4% of the variability ($R^2$). The Standardized Beta Weights indi-

cated that the Visual Perception Subtest had more predictive validity (*B*=.314) than the Motor Coordination Subtest (*B*=.23). The equation obtained was: Y = .227 VP + 71.89. It can be seen that the Motor Coordination Subtest did not appear in the equation, for its *p* value was not significant (.116). The correlation obtained between the Main Test and the Visual Perception Subtest was .41, while the correlation obtained between the Main Test and the Motor Coordination Subtest was .32.

*Evidence based on the relationship with other variables.* Three *t* tests were performed to assess differences according to gender (one for each test, Main Test and the two subtests; see Table 5). The Levene Test indicated that the assumption of homogeneity of variance was met for each test ($F = 1.454$, $p = .234$, Main Test; $F = .837$, $p = .365$, Visual Perception; $F = .373$, $p = .544$, Motor Coordination). Results indicated no significant differences between genders. The Cohen's *d* coefficient with Hedges correction for bias indicates that all effect sizes are small.

*T* tests were also performed to assess differences according to previous school experience (see Table 6). The homogeneity of variance assumption was met according to the Levene's test: $F = .006$, $p = .939$, Main Test; $F = 2.574$, $p = .115$, Visual Perception; and $F = .010$, $p = .922$, Motor Coordination. Results indicated that there was a significant difference in the Main Test between the performance of those who had and those who did not have previous school experience ($p= .032$). No significant differences were found for the Visual Perception subtest and the Motor Coordination subtest. It is noteworthy that only six children did not have previous school experience, while 44 had previous school experience. Therefore, these results need to be observed carefully. The Cohen's *d* coefficient with Hedges correction for bias indicates that the Effect size for the Main Test was large, while those of the other tests were small.

Another set of *t* tests were performed to assess differences according to socioeconomic status (see Table 7). The Levene Test indicated that the homogeneity of variance assumption was met for both subtests: Visual Perception ($F = .097$, $p = .756$) and Motor Coordination ($F = 2.693$, $p = .107$). However, it was not met for the Main Test ($F = 4.683$, $p = .035$). Therefore, the information provided by the *t* test was evaluated assuming equal variance for the subtests and different variance for the Main Test. The *t* test results indicated that there were significant differences for the Visual Perception subtest ($p = .016$). No significant differences were found for the Motor Coordination subtest and the Main Test. The

Table 5
*Descriptive statistics for each of the Beery VMI tests in regards to gender*

| | ***Feminine*** | | ***Masculine*** | | ***Independent samples t test results*** | | | |
|---|---|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | ***M*** | ***SD*** | ***t*** | ***df*** | ***p*** | ***ES*** |
| *Main Test* | 92.36 | 10.641 | 92.68 | 8.611 | -.117 | 48 | .907 | -.0325 |
| *Visual Perception* | 87.56 | 16.699 | 93.88 | 17.864 | -1.292 | 48 | .202 | -.3598 |
| *Motor Coordination* | 92.52 | 11.899 | 90.36 | 10.234 | .688 | 48 | .495 | .1916 |

*Notes*. *M* = mean; *SD* = standard deviation; ES = Effect size. Sample size for each group was *n* = 25.

Table 6
*Descriptive statistics for each of the Beery VMI tests in regards to previous school experience*

| | ***School experience previous to Kindergarten (n = 44)*** | | ***No school experience previous to Kindergarten (n = 6)*** | | ***Independent samples t test results*** | | | |
|---|---|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | ***M*** | ***SD*** | ***t*** | ***df*** | ***p*** | ***ES*** |
| *Main Test* | 91.45 | 9.232 | 100.33 | 9.136 | -2.212 | 48 | .032 | -.9478 |
| *Visual Perception* | 90.77 | 18.251 | 90.33 | 10.250 | .057 | 48 | .954 | .0246 |
| *Motor Coordination* | 91.91 | 10.776 | 88.00 | 13.387 | .811 | 48 | .421 | .3474 |

*Notes*. *M* = mean; *SD* = standard deviation; ES = Effect size.

Table 7
*Descriptive statistics for each of the Beery VMI tests in regards to socioeconomic status*

| | ***High*** | | ***Low*** | | ***Independent samples t test results*** | | | |
|---|---|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | ***M*** | ***SD*** | ***t*** | ***df*** | ***p*** | ***ES*** |
| *Main Test* | 93.84 | 7.476 | 91.20 | 11.310 | .974 | 41.61 | .336 | .2711 |
| *Visual Perception* | 96.56 | 16.363 | 84.88 | 16.732 | 2.495 | 48 | .016 | .6947 |
| *Motor Coordination* | 92.00 | 12.530 | 90.88 | 9.545 | .356 | 48 | .724 | .0990 |

*Notes*. *M* = mean; *SD* = standard deviation; ES = Effect size. Sample size for each group was *n* = 25.

Table 8
*One-way analysis of variance results*

| | ***Mother's educational level*** | | | ***Father's educational level*** | | |
|---|---|---|---|---|---|---|
| | ***F*** | ***p*** | ***ES*** | ***F*** | ***p*** | ***ES*** |
| *Main Test* | .163 | .921 | -.0529 | .777 | .513 | -.0136 |
| *Visual Perception* | 2.025 | .124 | .0579 | 1.429 | .246 | .0251 |
| *Motor Coordination* | 1.132 | .346 | .0078 | 1.235 | .308 | .0139 |

*Notes*. Between groups *df* = 3 and Within groups *df* = 46 for all tests.

Cohen's *d* coefficient with Hedges correction for bias indicates that the effect size for the Visual Perception subtest was moderate, while the other two tests obtained small effect size coefficients.

A One-way Analysis of Variance was performed to examine differences in children's performance when considering the parents' educational level. A total of six ANOVAs were performed: three using father's educational level and three using mother's educational level, one for each test (Main Test and the two subtests). The educational level consisted of four levels: less than high school, high school, from one to three years of college, and college education of four years or more. The homogeneity of variance assumption was met for all tests. Results indicated that there were no differences in performance according to parents' educational level (see Table 8). The omega squared coefficient indicated that all but one effect size was small. The effect size for the Visual Perception subtest using Mother's educational level as independent variable was moderate.

*Evidence based on test consequences.* Three one sample *z* tests (the Main Test and the two subtests) were performed (see Table 9). The average score found in each test was: 92.52 for the Main Test, 90.72 for the Visual Perception Subtest, and 91.44 for the Motor Coordination Subtest. No significant difference was found for any of the tests. Therefore, the average performance of the sample does not differ from the average (mean = 100; standard deviation =15; *n* = 192) reported by Beery and Beery (2004). The Cohen's *d* coefficient with Hedges correction for bias indicates that all effect sizes are moderate.

Table 9
Z test results

| | ***M*** | ***SD*** | ***z*** | ***ES*** |
|---|---|---|---|---|
| *Main Test* | 92.52 | 9.582 | -.0705 | -.5302 |
| *Visual Perception* | 90.72 | 17.409 | -.0875 | -.5960 |
| *Motor Coordination* | 91.44 | 11.038 | -.0807 | -.5975 |

*Notes*. *M* = mean; *SD* = standard deviation; ES = Effect size. Sample size *n* = 50.

In regards to the interviews conducted with the four professionals, all of them stated that the Beery VMI is not the only instrument they used. Also, they interpret the test's numeric results along with the clinical observations performed during the evaluation process.

> OT2--- Well, according to the test results I see where the child is... average, under average and where he or she stands in regards to children of his or her own age. There are times that even when the test places the child at the average, one as a professional knows that the child is a little at the borderline and there are areas that need to be improved.

> P1--- Tables give you final numbers, but we obviously analyze those numbers within the context of the observations.

Since all professionals emphasized the use of clinical observations, this emerged as a category during the data analysis. In general terms, the Occupational Therapists emphasized the observation of various motor aspects during the child's performance, while psychologists also mentioned another type of observation (academic aspects).

> OT1--- observations made along the way, like posture, the way the pencil is held, how the middle line is crossed, how we integrate the other hand to assist, how the information given is interpreted.

> P2--- observations regarding tracing, instructions, time he or she took to look at the model and reproduce it, how does he or she make adjustments to draw the pictures, movements, if he or she becomes anxious.

These professionals also emphasized the use of other tests as criteria or complements to the Beery VMI within an evaluation session. The OTs mentioned the use of the Test of Visual Motor Skills (TVMS) and the PeaBody Developmental Motor Scales. Both tests evaluate motor skills, including visual-motor integration skills. They mentioned that they look for other tests to which compare the results of the Beery VMI. For example, OT1 stated: "I use them to confirm or discard what one sees". The two psychologists stated that they use the Wechsler Intelligence Scale - Revised to complement the Beery VMI and their clinical observations.

When professionals were asked specifically about the decisions they make based on the Beery VMI, they indicated once more that it is not the only criterion they use:

> OT2--- As complement to other type of evaluations, because I determine at what level is his or her integration, coordination, and visual perception, and I see whether the child needs a specific intervention in those areas.

> P1--- the decisions I make based on the results are: first, if the child needs to be evaluated by an occupational therapist; second, sometimes it helps to convey a diagnostic that you are considering; and third, to establish a work plan.

Finally, as part of the interview, professionals were asked their opinion regarding the use of the normative data provided by the Beery VMI and its applicability to Puerto Rican children. This time, opinions were diverse. The first Psychologist talked in favor of its applicability:

> P1--- I understand that it is an instrument that in a simple and fast way gives us essential information that can be related to the child's performance in different areas: reading, writing, mathematics, and, eventually, in his or her academic achievement. Like I said before, it has not fail me yet, when I make an evaluation and find some difficulty in the child's visual-motor integration, visual perception, and motor coordination parents end up calling me again in two or three years.

On the other hand, OT2 and P2 indicated that, even when it worries them that the test is not normalized for Puerto Rican children, they do not think that this affects significantly the validity of its results, because the test does not entails the use of culturally biased vocabulary:

> OT2--- The ideal case will be that there was an instrument normalized for Puerto Rican children. However, in the field of occupational therapy there is no test normalized in PR. Therefore, since there is none, we have to work with what we have and complement it with other evaluations. Also, one has to see that it is not a vocabulary test that one has to translate. It is a test of motor skills where one gives the child instructions and the child follows them. I do not think that there shall be too much variability between the normalization in the US and in PR. It applies well to Puerto Rican children.

> P2--- I think that many psychologists are using tests that do not apply to the Puerto Rican population. I think that the problem we have, in general, not only with the VMI, is that we do not have norms for Puerto Rican children, and even if some people do not want to acknowledge it, we are a different culture. (However,) I would say that it could be applied (referring to the Beery VMI) under the assumption that it is not a verbal test.

OT1 was the one who indicated more concern in regards to the use of normative data and its applicability to Puerto Rican children:

> OT1--- Puerto Rican children display great variability. For example, the population with whom I am working now has been greatly exposed to pencils, crayons, tracing, drawing, blocks… But, populations with whom I had worked in other instances, well, have been from low income families. The latter are children that have never hold a pencil or scissors. I think that this test has been normalized with a population that had many opportunities or at least belong to a culture that promotes such opportunities.

Based on the information gathered, there is agreement in the information given by the professionals in regards to the use of other instruments and observations to interpret the results of the Beery VMI and the emphasis in that decisions are not solely based in the results of the Beery VMI. However, they posses different opinions regarding the applicability of the Beery VMI results to the children they evaluate.

## Discussion

Table 10 summarizes the findings of this study, classifying them into two categories: those that support and those that do not support the validity of the Beery VMI. Sources of evidence based on test content, response process, and test consequences are more supportive of the validity of the results of the Beery VMI. On the other hand, findings related to the sources of evidence based on the test's internal structure and the relationships with other variables are not consistent.

To some extent, it can be argue that Beery and Beery's (2004) claim regarding the "none bias" status of the test was sustained according to the findings associated with gender, parents' level of education, socioeconomic status, and previous school experience. An exception occurred with the results of the Visual Perception Subtest when compared on socioeconomic status and previous school experience. It is important to point out that during the analysis process associated with these findings it was not possible to use the Central Limit Theorem to assume that the distribution of sample means approximated the normal distribution. In addition, Hinkle, Wiersma, and Jurs (2003) advice that a small sample size could increase the chance of not finding enough evidence to reject the null hypothesis.

On the other hand, according to the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999), the most common way to provide evidence based on the relationship with other variables is through the administration of additional tests which measure similar constructs to those examined on the test being evaluated. Since, there are no other tests, similar to the Beery VMI, normalized for Puerto Rican children; it was decided to use some categorical variables to provide data related to this source of evidence. Nonetheless, P2 comments pointed out that, sometimes, the child's results in the Beery VMI are not consistent with the results of the Wechsler Intelligence Scale for Children (WISC). However, P1 did not make similar comments.

In regards to the results related to the source of evidence based on the internal structure of the test, the Pearson correlation coefficient hypothesis tests indicated that there were no significant differences between the correlation coefficients presented by Beery and Beery (2004) and the ones calculated with the results of the sample of this study. On the other hand, the results related to the comparison of the Alpha internal consistency coefficients indicate that those calculated with the sample results were significantly different for the Main Test and the Motor Coordination Subtest but not for the Visual Perception Subtest.

There is one important aspect we should look at regarding the findings of the Pearson coefficients. The hypothesis tests pointed out that the correlation coefficients calculated with the results of the sample are consistent with those presented by Beery and Beery (2004). We should recall that the authors of the test indicate that there should be a correlation, at least moderate, between the Beery VMI subtests and its Main Test of Visual Motor Integration. Nonetheless, Beery and Beery (2004) do not establish clearly what would be an adequate value to state that the relationship among the tests and its subtests is moderate. For example, according to Hinkle, Wiersma, and Jurs (2003) the Pearson correlation coefficients values calculated with the results of the sample of this study, as well as those calculated by Beery and Beery (2004) are considered low and positive. Considering this and the relatively low variance percentage (21%) on the Main Test that, according to the regression analysis, could be explained by the scores the children obtained on the Visual Perception and Motor Coordination Subtests; it is difficult to establish decisively whether these evidences support or not the validity of the Beery VMI results based on the sample of this study.

In general terms, although three of the five sources of evidence evaluated (test content, response process, and test consequences) pro-

Table 10
*Findings that support and do not support the validity of the Beery VMI results obtained with the sample of Puerto Rican children*

| *Sources of evidence* | *Support validity of results* | *Do not support validity of results* |
|---|---|---|
| *A. Test content* | • Interviewed professionals indicated they use the Beery VMI for the test established purposes.<br>• Interviewed professionals evaluated positively the Beery VMI content, pointing out the following aspects: organization and presentation of the items, way in which the child is asked to work, and openness the children show toward the test. | • One of the interviewed professionals expressed concern in regards to the validity of the Beery VMI results, especially, when administered to preschool children. She said, she believed that the content of the tests' first items consist of draws that, usually, preschoolers have over-learned. |
| *B. Response process* | • During the cognitive interviews, when observing the figures, children named all the ones they know. If they did not know the name of the figure, they began to associate the forms with other things they did know.<br>• Children demonstrated comprehension of the directions resulting from the back translation process.<br>• Some children indicated that, as the test progressed, the figures were harder.<br>• It was not considered necessary to modify the administration process of the test, neither its directions. | |
| *C. Test's internal structure* | • There were no significant differences in Pearson correlation coefficients calculated on this study and the ones calculated by Beery and Beery (2004) based on the results of the normative sample. | • Only the Alpha reliability coefficient of the Visual Perception Subtest indicated a high level of internal consistency.<br>• The Cronbach's Alpha coefficient hypothesis test indicated significant difference between the ones presented on the test's manual and the ones calculated with the results of the sample for the Main Test and the Motor Coordination Subtest. |
| *D. Relationship with other variables* | • There was no significant difference between girls' and boys' performance.<br>• There were no significant differences among children's performance according to the educational level of their parents.<br>• On the Main Test and the Motor Coordination Subtest, there was no significant difference in regards to previous preschool experience and socioeconomic status. | • There was significant difference between the performance of this study sample and the normative sample on the Main Test in regards to previous school experience.<br>• There was significant difference between the performance of this study sample and the normative sample on the Visual Perception Subtest in regards to socioeconomic status. |
| *E. Tests' consequences* | • There was no significant difference between the average performance of this study sample and the one of the normative sample.<br>• The interviewed professionals pointed out that they used to interpret Beery VMI results considering also children results on other tests and their own clinical observations.<br>• According to the professionals, the Beery VMI contributes to the decision making in regards to the following aspects: child's difficulty level, pertinence of treatment, possibility of evaluation with other professionals, and recommendations for home follow up.<br>• Three of the interviewed professionals expressed positively in regards to the applicability of the Beery VMI results to Puerto Rican children. | • One of the interviewed professionals questioned the applicability of the Beery VMI norms to the Puerto Rican children. She argued that the test has been normalized with children whose culture has promoted their exposition to fine motor development activities (e.g. coloring, cutting), different from the Puerto Rican children culture, especially those of lower socioeconomic status. |

vided findings that, partially, sustain the validity of the Beery VMI results, it is not possible to establish that the results obtained by the sample of Puerto Rican children definitely confirm the validity of the test when considering all the sources of validity evidence established on the Standards for Educational and Psychological Testing [AERA, APA, & NCME, 1999]).

## Clinical and educational implications

Although, it is not possible to establish conclusively the validity of the results of the Beery VMI based on the findings of this research, it is neither possible to discard it. Therefore, the strategy used by the clinicians in regards to the use of their clinical observations and other tests to complement the Beery VMI results during the evaluation processes is considered adequate, since it allows for the triangulation of the information regarding the child's performance.

On the other hand, the lack of instruments normalized with Puerto Rican children, corroborated through this study, points at the need for professional education programs to promote research projects involving tests normalizations in Puerto Rico. This challenges professional education programs to work in the establishment of strategies to reinforce the research component of their curriculums to promote that their graduates feel prepared to perform, not just as clinicians, but also as researchers.

Based on the results of this study, the possibility of the use of the Beery VMI in schools to contribute to the early identification of children with needs related to the skills evaluated by the test, should not be discarded. The need for early identification strategies in Puerto Rico was established by Bonilla-Landrón (2002) in her study related to preschoolers with and without developmental needs. However, this decision would need to be supported with more research. In her study, Bonilla-Landrón (2002) also emphasized the need to increase teachers' knowledge of the skills needed to develop better handwriting, especially in preschool. This is consistent with teachers' concerns found during the schools visits performed in this study. Teachers indicated that they have had children they thought had certain needs, but they did not feel capable to distinguish whether children difficulties were due to immaturity or to skills deficiencies. This area deserves more research, since it could indicate the need to include this aspect in preschool teachers' preparation programs. After all, teachers (not therapists) have greater opportunities to observe children and identify their needs.

## Contributions and limitations of the study

This study makes a contribution to the Occupational Therapists and Psychology fields in Puerto Rico, since the literature reviewed did not find studies which examined the validity aspects of the results of the Beery VMI for a sample of Puerto Rican children. According to the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999), the validation of an instrument is responsibility of its author as well as the professional who uses it. On the other hand, the importance of Evidenced Based Practice (EBP) has been emphasized in OT and in other health related fields. Sackett, Rosenberg, Gray, Haynes, and Richardson (1996, in Christiansen & Lou, 2001) stated that EBP involves the conscious and explicit use of the best and most current evidence during the decision process in regards to the care of our patients. Therefore, by providing evidence about the sources of validity of the Beery VMI, this study is contributing to the EBP of the Occupational Therapists and Psychologists in Puerto Rico. In addition, the literature reviewed did not reveal research efforts in which the professionals' opinions about their use of the test (evidence based on test consequences) were used as evidence related to the validity of the test's results.

On the other hand, there are some aspects that limit the generalization of this study's findings. The sample consisted of two groups of Kindergarteners from two different schools (50 children) and four professionals selected through a convenience sampling method. The small sample size of children limits the ability to perform inferential statistics. The sampling method (convenience sample) is another aspect that limits the generalization of findings. These factors represent a threat to the external and statistical validity of the results. However, this study provides a road map on how to gather validity evidences for a larger and more rigorous study regarding the VMI, as well as for researchers and practitioners wanting to collect validity evidence when using standardized tests. Additional limitations of the study are related to the maturity of the children. The Beery VMI was administered to Kindergarteners in the second semester of the academic year; thus, the maturity level of these children is not the same to the one of Kindergarteners on their first semester.

## Future directions

Studies should be conducted to replicate this research using larger samples and other sampling methods to allow for the generalization of the findings. They may look for validity aspects of the Beery VMI

using groups from different ages and verify if the test is able to differentiate those groups.

Another study can use a qualitative approach to analyze professionals' documentation in order to examine if recommendations made based on the Beery VMI results are consistent with its intended purposes. A third study can survey preschool teachers to look at aspects like the criteria they use to identify skill difficulties in their children or their perceived ability to identify children's needs. One last study can focus on the development of a standardized observation guide for the professionals who use the Beery VMI to direct them uniformly in the observation of the subject's performance quality and process during the test.

## REFERENCES

Allen, M., & Yen, W. (2002). *Introduction to measurement theory* (2nd ed.). Long Grove, IL: Waveland Press.

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Beery, K., & Beery, N. (2004). *The Beery-Buktenica Developmental Test of Visual-Motor Integration: Administration, scoring and teaching manual* (5th ed.). Minneapolis, MN: NCS Pearson.

Bonilla-Landrón, M. (2002). *La lectoescritura en preescolares con y sin rezago en el desarrollo y el aprendizaje: Discursos y prácticas de dos maestras de Kindergarten.* Unpublished master's thesis, Universidad de Puerto Rico, Río Piedras, Puerto Rico.

Braverman, M. T., & Slater, J. K. (1996). *Advances in survey research: New directions for evaluation.* San Francisco, CA: Jossey-Bass.

Case-Smith, J. (2002). Effectiveness of school-based occupational therapy intervention on handwriting. *American Journal of Occupational Therapy, 56* (1), 17-25.

Christiansen, C., & Lou, J. (2001). Ethical considerations related to evidence-based practice. *American Journal of Occupational Therapy, 55* (3), 354-349.

Creswell, J. (2005). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (2nd ed.). Upper Saddle River, NJ: Merill-Prentice Hall.

Cornhill, H., & Case-Smith, J. (1996). Factors that relate to good and poor handwriting. *American Journal of Occupational Therapy, 50*(9), 732-739.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Crano, W., & Brewer, M. (1973). *Principles of research in social psychology*. New York: McGraw-Hill.

Daly, C., Kelley, T., & Krauss, A. (2003). Relationship between visual motor integration and handwriting skills of children in Kindergarten: A modified replication study. *American Journal of Occupational Therapy, 57*(4), 459-462.

Fowler, F. J. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Frey, P., & Pinelli, B. Jr. (1991). Visual discrimination and visuomotor integration among two classes of Brazilian children. *Perceptual and motor skills, 72*, 847-850.

Gorsuch, R. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hinkle, D., Wiersma, W., & Jurs, S. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston, MA: Houghton Mifflin Company.

Hsu, H. (1997). *Performance of primary school children in Taiwan on the Beery Developmental Test of Visual-Motor Integration (China)*. Unpublished master's thesis, University of Southern California, California, United States.

Institute of Phonetic Sciences, Amsterdam. (1996). Two correlation coefficients [Web service]. Retrieved August 29, 2006 from http://www.fon.hum.uva.nl/Service/Statistics/Two_Correlations.html

Kulp, M., & Sortor, J. (2003). Clinical value of the Beery Visual-Motor Integration Supplemental Tests of visual perception and motor coordination. *Optometry and Vision Science, 80*(4), 312-315.

Marr, D., Windsor, M., & Cermak, S. (2001). *Handwriting readiness: Locatives and visualmotor skills in the Kindergarten year.* Utica, NY: Early Childhood Research & Practice: An Internet Journal on the Development, Care and Education of Young Children. (ERIC Document Reproduction Service No. ED452998)

Marriott, G. (2000). *Developmental differences in visual-motor integration in children from 4 to 6- years-old on the Kindergarten Diagnostic Instrument-Second edition.* Unpublished doctoral dissertation, Texas Woman's University, Texas, United States.

McHale, K., & Cermark, S. (1992). Fine motor activities in elementary school: Preliminary findings and provisional implications for children with fine motor problems. *American Journal of Occupational Therapy, 46*(10), 898-903.

McMillan, J. (2004). *Educational research: Fundamentals for the consumer* (4th ed.). Boston, MA: Pearson Education.

Messick, S. (1989). Meaning and values in test validation: The sciences and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

Portney, L., & Watkins, M. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). Upper Saddle River, NJ: Prentice Hall Health.

Richardson, P. (2001). Use of standardized test in pediatric practice. In J. Case-Smith (Ed.), *Occupational therapy for children* (pp. 217-245). Saint Louis, MO: Elsevier Science.

Stevens, J. (2001). *Applied multivariate statistics for the social sciences* (4ht ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Allyn & Bacon.

Tseng, M., & Murray, E. (1994). Differences in perceptual-motor measures in children with good and poor handwriting. *Occupational Therapy Journal of Research, 14* (1), 19-36.

Verdonck, M. C., & Henneberg, M. (1997). Manual dexterity of South African children growing in contrasting socioeconomic conditions. *American Journal of Occupational Therapy, 51*, 303-306.

Weintraub, N., & Graham, S. (2000). The contribution of gender, orthographic, finger function, and visual-motor processes to the prediction on handwriting status. *Occupational Therapy Journal of Research, 20*(2), 321-34.

Este artículo se recibió en la Redacción de *Pedagogía* en febrero de 2009 y se aceptó para su publicación en marzo del mismo año.