In a comprehensive assessment of upper extremity function, dexterity is an important component that occupational therapists must consider (Kohlmeyer, 1998; Mathiowetz & Bass-Haugen, 1995). Dexterity has been defined as “the fine, voluntary movements used to manipulate small objects during a specific task, as measured by the time to complete the task” (Backman, Cork, Gibson, & Parsons, 1992, p. 208). Chan (2000) defined it as “the skillful and controlled manipulation of a tool or an object by the fingers” (p. 537). Dexterity is considered essential for successful performance of tasks of daily living, work, school, play, and leisure (Exner, 1990).

One of the most commonly used tools for assessing dexterity is the Nine Hole Peg Test. This test was originally introduced in 1971 as part of a study on strength and dexterity (Kellor, Frost, Silberberg, Iversen, & Cummings). This study gave approximate dimensions for the peg board and the pegs and general procedures for administration. In 1985, Mathiowetz, Weber, Kashman, and Volland published a study that provided more detailed instructions for administration, evaluated the interrater reliability and test–retest reliability, and established adult norms. This study also provided instructions for construction of the Nine Hole Peg Test, however, none was ever produced commercially using these specifications.

According to a recent survey (Davis, Kayser, Matlin, Mower, & Tadano, 1999) of the multiple versions commercially available today, the Nine Hole Peg Tests produced by Smith & Nephew Rehabilitation Division version of the Nine Hole Peg Test, and to establish new adult norms for the Nine Hole Peg Test for finger dexterity utilizing this particular version. Two of the researchers established interrater and test–retest reliability by evaluating 25 occupational therapy student volunteers. Seven hundred and three subjects, ranging in age from 21 to 71+ years, were tested to establish norms, using the standard protocol. Results showed high interrater reliability and only moderate test–retest reliability. Scores obtained by using the commercially available version were not statistically different from previously published norms (Mathiowetz, Weber, Kashman, & Volland, 1985). This study supports the original norms and further assists occupational therapists to evaluate dexterity accurately.
should be used with caution for the newer commercially available versions. The survey by Davis et al. (1999) found, however, that a large number of therapists do use the Mathiowetz et al. published norms with the commercially produced versions of the test and that there is a strong interest in norms being produced for the newer versions.

Considering this information, we designed a study to establish new norms for one of the commercially available versions. In comparing the original Mathiowetz et al. (1985) version with the two most commonly used commercial versions, the Smith & Nephew version is the closest in terms of dimensions. The only differences between the two tests are that the commercial version is made from plastic instead of wood, and that the container for the pegs is attached or built in, which doubles the width of the instrument compared to the original. Another change made in the commercial version is that the peg container or cup is round instead of square. In the Mathiowetz et al. study, one problem cited was that some subjects had difficulty picking up pegs from the corners of the square cup.

The purpose of our study was to evaluate the interrater and test–retest reliability of the most widely-used, commercially available version of the Nine Hole Peg Test and to establish new adult norms by gender using this particular version. Therefore, we chose to duplicate the study by Mathiowetz et al. (1985) that established the currently used norms.

Methods
As in the original study by Mathiowetz et al. (1985), this study was composed of two parts: a preliminary reliability study and a normative study. In the first part of the study, two faculty members evaluated 25 occupational therapy student volunteers to establish interrater and test–retest reliability data. In the second part of the study, participants were recruited from the community for the establishment of normative data.

Reliability Study
The 25 participants used for this portion of the study were occupational therapy student volunteers. As in the Mathiowetz et al. (1985) study, participation in the study was limited to individuals without history of neuromuscular or orthopedic dysfunction that would significantly affect dexterity. Once the students gave informed consent, they were given instructions per the protocol outlined in the original study and were allowed a practice trial prior to the recorded trial. The participants were evaluated simultaneously by the two faculty researchers who independently timed and recorded the Nine Hole Peg Test scores to determine interrater reliability. All participants were then reevaluated by the same researchers 1 week later under the same standardized instructions to determine test–retest reliability.

Normative Study
Seven hundred and three participants were included in this part of the study (314 males and 389 females; Table 1 & Table 2). Participant ages ranged from 21–70+ years. Following the same format of the original study (Mathiowetz et al., 1985), age groups were preestablished for data representation purposes as: 21–25, 26–30, 31–35, 36–40, 41–45, 46–50, 51–55, 56–60, 61–65, 66–70, and 70+. Participants were recruited from a variety of community and institutional settings. After it was determined the participants met the specified criteria and had read and agreed with the approved consent form, each one was briefly interviewed prior to the test and the following information obtained: age, gender, and hand dominance.

Test Procedures and Instructions
The Nine Hole Peg Test used for data collection in the present study consists of a plastic console with a shallow round dish to contain the pegs on one end of the console and the nine-hole peg-board on the opposite end. Each participant was asked to center the pegboard directly in front of him or her, oriented such that the shallow dish was on the participant’s dominant hand side and the peg holes on the nondominant side. Instructions for the test were given as per the standard instructions (Mathiowetz et al., 1985) along with a brief demonstration. Participants were given the opportunity for a brief practice test prior to the actual test and were then tested using their dominant hand first, followed by their nondominant hand.

The tests were timed, with a stopwatch, from the moment the participant touched the first peg until the moment the last peg hit the dish. The test was then repeated for the nondominant hand using the same testing method, with the pegboard rotated such that the dish was in front of the nondominant hand. All participants were tested using this procedure. In the event that the participant dropped a peg or the trial was interrupted in any way, the evaluator cued the participant to stop and a new trial was initiated.

Results
Reliability Study
In the reliability study, both the interrater and test–retest reliability values were calculated using the Pearson correlation coefficient. A very high interrater reliability was obtained for both the right and left hands ($r = 0.994$ and $r = 0.993$, respectively). The test–retest reliability coefficient was low to moderate for both the right and left hand ($r = 0.459$ and $r = 0.442$, respectively).

In addition, a repeated measures analysis of variation (ANOVA) was done. The repeated factor was “trial” (trial 1 vs. trial 2) and the second factor was “rater” (rater 1 vs. rater 2). This design enabled the authors to test for differences in means between trial 1 and trial 2, differences between the two raters, and the interaction between trials and raters. For the right hand, there was no statistically significant difference for the interaction between rater and time ($F(1,24) = 1.916$, $p = .179$) nor between the raters ($F(1,24) = 2.350$, $p = .138$). There was a significant difference between trial 1 and trial 2 ($F(1,24) = 7.643$, $p = .011$). For the left hand, however, there was a statistically significant interaction between rater and time ($F(1,24) = 16.099$, $p = .001$), and, like the right hand, a significant difference between the trials ($F(1,24) = 5.284$, $p = .011$). The left hand also demonstrated no significant difference between the raters ($F(1,24) = 2.554$, $p = .123$).

To examine the significant interaction further, paired $t$ tests were used to test the
simple main effects. Two of the simple main effects (raters and trials) were statistically significant \( (p < 0.05) \): difference between the raters at trial 1, and differences between the trials for the second rater. The data indicate that the trial 1 mean for the second rater was significantly greater than the other means. In addition, the main effect for trials was significant; the trial 1 mean for both raters was significantly greater than the trial 2 mean.

**Normative Study**

As in the original normative study for the Nine Hole Peg Test, and other assessment tools, such as the dynamometer for grip testing, data are presented by gender. The average performance and standard deviation values for male participants on the Nine Hole Peg Test are reported in Table 1, and female participants, in Table 2.

For males, the 21–25 age group demonstrated the highest performance (lowest scores), and the 71+ age group demonstrated the lowest performance (highest scores). For females, the 26–30 age group showed the highest performance (lowest scores), while the 71+ age group showed the lowest performance (highest scores). As expected, a high correlation was found between performance on the Nine Hole Peg Test and age (males: right hand \( r = 0.908 \), left hand \( r = 0.918 \); females: right hand \( r = 0.890 \), left hand \( r = 0.896 \)).

The standard deviation values for both males and females, up to age 70, ranged from 1.55 to 3.97. Standard deviation increased with increasing age. In the 71+ age group, the standard deviation was dramatically higher as compared to the other age groups.

When comparing the data in Table 1 and Table 2, the average female performed slightly better (lower score) on the Nine Hole Peg Test as compared to the average male subject. Also, the standard deviation for all female participants is lower as compared to all male subjects, indicating a lower variability in the scores of the female participants.

For both the males and females (all age groups), the scores for the right hand are slightly lower than the left hand, indicating slightly better performance using the right hand. This finding is probably related to the hand dominance of the participants. Approximately 90% of male participants and 93% of female participants were right-hand dominant. On average, right-hand dominant subjects scored better (lower scores) on their right hand as compared to their left hand. Clearly, hand dominance plays a role in the performance of the right hand versus the left hand. However, since left dominant participants made up less than 10% of the total sample, normative data are presented as right and left hands, combining the right and left dominant subjects.

The results of this study were compared to the results of the Mathiowetz et al. (1985) study. An independent \( t \) test was performed for each age group for both males and females, up to age 70. To protect against inflated family-wise error rates, Bonferroni’s procedure was used to adjust the alpha for each test \( (.05/11 \) comparison, \( p < .005 \)). For each age group up to 70 years, the \( t \)-test results indicate no significant differences in the average scores between this study and the previous study. Due to classification differences and the disparity between the numbers of participants 71 years and older, a comparison between the two studies for this particular age group was not performed.

**Discussion**

The interrater reliability of the Nine Hole Peg Test was found to be very high (right: \( r = 0.984 \), left: \( r = 0.993 \)). This value shows that two different test administrators will score the test almost exactly the same. As in the Mathiowetz et al. (1985) study, the test–retest reliability showed a moderate correlation in the scores of the subjects, which demonstrates poor reliability. Although the results of the repeated measures ANOVA for the interaction between rater and time showed a statistically significant difference for the left hand, this difference is probably not clinically significant, representing only a difference of 2/10 of a second. The significant difference between
trial 1 and trial 2 for both hands does suggest a possible practice effect. It has been suggested in previous research that data from multiple trials may reduce the practice effect. Mathiowetz et al. (1985) recommended the use of multiple trials for future research to improve test–retest reliability. In fact, Davis et al. (1997) demonstrated that the average of three trials produced a higher test–retest reliability (right $r = .815$, left $r = .895$) than a single trial. For this study, however, the authors chose not to include multiple trials in order to replicate the Mathiowetz et al. (1985) study as closely as possible in order to make direct comparisons between the two sets of norms. Based on the results obtained by Davis et al. (1997), it seems that more valid norms could be established with the use of three trials. This is a limitation of our current study. One of the questions the authors wanted to answer with this study was, “Are the normative scores for the Nine Hole Peg Test calculated in the 1985 study by Mathiowetz et al. applicable to other Nine Hole Peg Test versions?” In comparing findings from this study of the Smith & Nephew version to the previous study of Mathiowetz et al. (1985), no statistically significant differences were found in the normative scores for each age range. The results of this comparison demonstrate that the values Mathiowetz et al. (1985) obtained are applicable to this commercially available version of the Nine Hole Peg Test. This finding is in direct contradiction to the findings of Davis et al. (1997). Several possible explanations for this inconsistency exist. First, the Davis et al. (1997) study included only a total of 29 participants, with the number of participants in each age group ranging from 1 to 7. The small number of participants in each age range calls into question the reliability of their results. Second, they tested their participants on the older homemade version of the test, as well as the Smith & Nephew version. These two sets of scores were the data they used to make correlations, whereas we compared the norms we obtained with the norms obtained by Mathiowetz et al. Therefore, the results of our study and the Davis et al. (1997) study cannot readily be compared in regard to the transferability of the scores between the older version and the commercially available version. The Davis et al. (1997) study does not present enough data to draw definite conclusions. However, due to the poor test–retest reliability obtained in both the Mathiowetz et al. study and our current study, reliability of these norms should be questioned.

The high interrater reliability has important treatment implications for occupational therapy. For example, in client evaluation, two different therapists could administer the test to the same client and obtain consistent results. In the clinical setting, the occupational therapist could use multiple trials for each administration of the test to increase test–retest reliability.

With so many different versions of the Nine Hole Peg Test available to occupational therapists, it was important to compare the results obtained in this study with the results obtained by Mathiowetz et al. (1985). Obtaining norms for this commercially available version was a logical step since it is one of the most widely used versions. Occupational therapists using the Smith & Nephew version of the test can use either the scores obtained by Mathiowetz et al. or those obtained in this study.

Future research studies regarding the Nine Hole Peg Test should establish a new set of norms utilizing the average of three trials. Occupational therapists should follow the instructions outlined in the original study (Mathiowetz et al., 1985) when using these normative data as a comparison for their clients, but consider three trials for increased validity.

### Acknowledgment

The authors would like to thank Douglas Murphy, PhD, for his invaluable knowledge and help with the statistical work for this project.

### References


