A study of inter-instrument reliability and concurrent validity between the Baseline Digital Dynamometer and the Jamar Hydraulic Dynamometer

Heather A. Deaton

Follow this and additional works at: http://utdr.utoledo.edu/graduate-projects
A Study of Inter-instrument Reliability and Concurrent Validity Between the Baseline® Digital Dynamometer and the Jamar® Hydraulic Dynamometer

Heather A. Deaton, B.S.

Research Advisor: Julie Jepsen Thomas, Ph.D., OTR/L, FAOTA

Occupational Therapy Doctorate Program
Department of Rehabilitation Sciences
The University of Toledo

May 2014

Note: This scholarly project reflects individualized, original research conducted in partial fulfillment of the requirements for the Occupational Therapy Doctorate Program, The University of Toledo.
Abstract

The purpose of this study was to investigate the inter-instrument reliability and concurrent validity of the Baseline digital dynamometer. Grip strength testing was conducted on 102 healthy male and female participants between the age of 18-50 using both the Baseline digital dynamometer and the Jamar hydraulic dynamometer in a randomized, crossover design. Data were analyzed using t-tests for paired measures and a Bland-Altman analysis. The instruments were highly correlated (measuring the same construct), however, there were statistically significant differences between the measurements of the two instruments with the Jamar measuring higher on average (p=.011). The Bland-Altman analysis indicated a bias (difference between the means) of 1.79 lbs. The 95% limits of agreement were between -11.87 and +15.46 lbs. This means that the grip strength measurements can vary up to 27.32 lbs. between the instruments. This study’s results indicate that the Baseline digital dynamometer does not have strong concurrent validity and therefore, therapists should not use these two instruments interchangeably with the same client. It is not valid and reliable to compare numbers obtained with the Baseline digital dynamometer to the norms, which were established with the Jamar hydraulic dynamometer as these two instruments do not consistently measure the same. Future research should establish normative data using the Baseline digital dynamometer so that there is a reliable and valid set of reference values to be used with the Baseline digital dynamometer.
A Study of Inter-instrument Reliability and Concurrent Validity Between the Baseline® Digital Dynamometer and the Jamar® Hydraulic Dynamometer

Occupational therapists use grip strength measurements to evaluate one aspect of hand functioning. Grip strength can be measured with a variety of instruments (Svens & Lee, 2005). Before using an instrument for measurement it should first be determined to be a valid and reliable measurement tool (Fess, 1986). The Jamar hydraulic dynamometer is the most commonly used grip strength assessment instrument and was used to establish normative grip strength data (Mathiowetz, Kashman, Volland, Weber, Dowe, & Rogers, 1985). However, digital dynamometers are becoming more popular in clinical settings because of their ease of use (Shechtman, Gestewitz, & Kimble, 2005). Past research has shown that not all dynamometers can be used interchangeably with each other nor with established norms (Flood-Joy & Mathiowetz, 1987; Svens & Lee, 2005; Allen & Barnett, 2011). To this end, the current study will examine whether there are differences in grip strength measurements between the Baseline digital dynamometer and the Jamar hydraulic dynamometer. In order to understand the present study, a review of the literature outlining the importance of reliability and validity of measurement tools will be provided. Past research on comparing grip strength measurement tools will be discussed followed by a description of the present study.

Why is grip strength measured?

The ability to grip an object is an important hand function required in many daily routines. Rice, Leonard, and Carter (1998) found that in a normal college age population, it requires 2-14 pounds of grip force to open common household containers and spray bottles. Shiffman (1992) concluded that decreases in hand grip strength are correlated with a decreased ability to perform functional tasks. This is important for occupational therapists because research
has shown that grip strength decreases with age (Shiffman, 1992; Ranganathan, Siemionon, Sahgal, & Yue, 2001). There is a curvilinear relationship between grip strength and age, with grip strength peaking between 20-50 years of age (Mathiowetz, Kashman, Volland, Weber, Dowe, & Rogers, 1985). Hand strength begins decreasing around the age of 50 and significant decreases occur around 70 to 80 years of age (Horowitz, Tollin, & Cassidy, 1997).

Hand grip strength is a long-standing measure for occupational therapy. The purpose of grip strength evaluation is to make strength comparisons relative to normative data, monitor individual progress, and collect information about the effectiveness of a rehabilitation program (Allen & Barnett, 2011; Mathiowetz, Vizenor, & Melander, 2000; Svens & Lee, 2005). Grip strength is also used to describe the relationship of health factors and functional status, including limitations in occupations of daily living and disability (Bohannon, 2008).

**How is grip strength measured?**

There are a variety of grip strength measurement devices that are used in clinical practice and research. These include hydraulic dynamometers, pneumatic instruments, strain gauges, and work evaluation devices with attachments (Svens & Lee, 2005). A study conducted by the California Medical Association committee and headed by Kirkpatrick (1956) determined that the hydraulic dynamometers provided the most accurate measurement of grip strength, when compared to the pneumatic instrument and a spring dynamometer. Dynamometers assess grip strength by measuring the force exerted by the hand during a maximal isometric contraction (Svens & Lee, 2005). There are a variety of dynamometers available for evaluating grip strength. However, not all instruments have been shown to be valid and reliable. Fess (1986) stated that an instrument’s reliability and validity has a direct effect on the quality of information that is obtained.
What is reliability and validity?

Reliability refers to whether a given instrument provides stable and consistent results (Kielhofner, 2006; Shechtman, Gestewitz, & Kimble, 2005). Reliability can refer to consistency between multiple raters (inter-rater reliability), stability over repeated measures (test-retest reliability), and the consistency between instruments (inter-instrument reliability) (Fess, 1986). Reliability is expressed as a correlation coefficient and standard error of measurement (Fess, 1986). Validity refers to whether an instrument measures what it is intended to measure (Kielhofner, 2006; Shechtman, Gestewitz, & Kimble, 2005). In order to have validity an instrument must first demonstrate reliability (Fess, 1986). The development of new models of dynamometers makes criterion-related validity and inter-instrument reliability of particular interest. Criterion-related validity or concurrent validity involves demonstrating a correlation between the measurement of interest and another instrument that has been shown to be accurate – the established “gold standard” (Kielhofner, 2006; Shechtman, Gestewitz, & Kimble, 2005). Furthermore, inter-instrument reliability is the extent to which different instruments measure equivalently under the same conditions.

Researchers have determined that the hydraulic dynamometer demonstrates high reliability and validity, in particular the Jamar hydraulic dynamometer (Bellace, Healy, Byron, & Hohman, 2000; Flood-Joy & Mathiowetz, 1987; Hamilton, Balnave, & Adams, 1994; Mathiowetz, Weber, Volland, & Kashman, 1984; Peolsson, Hedlund, & Oberg, 2001). In order to further improve the reliability and validity of the hydraulic dynamometer, The American Society of Hand Therapists (ASHT) has adopted a standardized positioning that should be used while taking grip strength measurements (Fess, 1992). They have also determined that when
using the Jamar hydraulic dynamometer, measurements should be taken in the second handle position (Fess, 1992) in order to get the most reliable and valid results.

**Perceived advantages and disadvantages of digital versus hydraulic**

In recent years, dynamometers with digital display readouts are being used in clinical settings. “The benefits of using a computerized dynamometer include the digital readout, which increases the inter-rater reliability; increased precision, which allows reading of the force scores to the closest pound as opposed to the five pound increments on the Jamar dynamometer; and the ability to electronically store and retrieve grip strength scores, which makes it convenient and time-saving” (Shechtman, Gestewitz, & Kimble, 2005, p. 343). However, the normative data for hand grip strength was established using the Jamar hydraulic dynamometer (Mathiowetz, Kashman, Volland, Weber, Dowe, & Rogers, 1985). Empirical evidence supporting the reliability and validity of digital dynamometers is limited.

**Interchangeability of grip strength instruments**

Researchers have established inter-instrument reliability and criterion-related validity for several models of dynamometers. Mathiowetz (2002) compared the Roylan and Jamar hydraulic dynamometers. Healthy subjects were recruited and included a convenient sample of 30 males and 30 females, age 20 to 50 years of age, right- or left-hand dominant. Paired-data t-tests indicated no significant difference between the two dynamometers. The intraclass correlation coefficients ranged from 0.90 to 0.97. With these data they were able to conclude that the Roylan hydraulic dynamometer showed acceptable concurrent validity and excellent inter-instrument reliability (2002).

Mathiowetz, Vizenor, and Melander (2000) compared the Baseline hydraulic dynamometer and Jamar hydraulic dynamometer in terms of inter-instrument reliability and
concurrent validity. The right hand grip strength of 80 healthy subjects was measured with both instruments. The authors found that there were no statistically significant differences for males and there was a 3.4% difference in grip strength scores for females. Since this was less than the 5.6% standard error of measure for grip strength measurements the authors determine that the difference was not practically significant. Intraclass correlation coefficients (ICCs), which is measured to determine interchangeability, were above .9994 supporting inter-instrument reliability. The authors concluded that the two dynamometers measure equivalently for practical purposes. Therefore, it would be acceptable to use the Baseline hydraulic dynamometer with the currently established grip strength normative data (2000). However, it should be noted that t-tests showed significant differences between the two dynamometers for the female grip strength scores.

In an initial study Flood-Joy and Mathiowetz found significant differences between two different models of the Jamar hydraulic dynamometer. The authors discovered that there were problems with the calibration of one of the dynamometers, so they completed a replication study comparing three versions of the Jamar hydraulic dynamometer. In this replication study, there were 30 subjects age 20 to 60 years. The results indicated that there were significant (p<0.01) differences in the strength readings among the three dynamometers. Version C was measuring approximately 6-11 pounds higher than Version A and Version B. The authors concluded that not all models of the Jamar hydraulic dynamometer read equivalently. Therefore, they stated that it is essential to use the same dynamometer for pre and post-testing clients (1987).

Svens and Lee compared a computer-connected GripTrack dynamometer to the Jamar hydraulic dynamometer to determine intra- and inter-instrument reliability. Forty-six female health care workers between the ages of 21 and 40 participated in the study. They found that the
Jamar hydraulic dynamometer recorded higher measurements than the GripTrack. Data showed that strength reading could vary up to 10.23 kg between the instruments. Their ICCs did not exceed 0.90 indicating that the two dynamometers can differ in grip strength readings between 2.22 kg and 3.28 kg 95% of the time. The authors concluded that the scores were not interchangeable between the two measurement tools and that the normative data should not be used with scores recorded from the GripTrack system (2005).

Allen and Barnett conducted a comparison study between the Biometrics E-link dynamometer (another computer connected system) and the Jamar hydraulic dynamometer. Participants included a convenient sample of 42 female and seven male university students age 18 to 25. Data showed that there was excellent concurrent validity for both hands (ICC= 0.986 for right hand and 0.983 for left hand). From these data the authors concluded that the Biometrics evaluation system is measuring the same construct as the Jamar hydraulic dynamometer. However, the right hand grip strength scores were higher for the Biometrics evaluation system. This difference was statistically significant (p< 0.05), indicating that therapists should exercise caution when interchanging instruments (2011).

Massy-Westropp, Ahern, and Hearn (2004) compared the electronic Grippit to the Jamar hydraulic dynamometer. This study had a large sample of 476 healthy males and females age 18-97 years old. Bland-Altman analysis of the mean differences between the two instruments revealed a bias (mean difference) of 22 Newtons and limits of agreement of -86 to 129 Newtons, which indicates that grip measurements may vary by up to 215 Newtons between the instruments. The authors concluded that the limits of agreement between the two dynamometers are too large to permit the instruments to be interchangeable in the clinic.
Data testing the criterion-related validity of a digital dynamometer to the Jamar hydraulic dynamometer is limited to two studies. Shechtman, Gestewitz, and Kimble tested the concurrent validity of the DynEx digital dynamometer. In this study they tested 100 healthy subjects between the ages of 20 and 40 years. Each person’s hand strength was measured on both hands with both dynamometers in a randomized order. They found that the Jamar hydraulic dynamometer measured forces that were consistently 2.6 to 4.9 lb. higher than the DynEx digital dynamometer (4.0-9.5%). These differences were statistically significant however they still showed excellent concurrent validity (r>0.98). They concluded that clinicians can use the DynEx digital dynamometer at the tested position knowing that it is measuring the same construct as the Jamar dynamometer but strength values cannot be interchanged. Therefore, the same dynamometer should be used for each therapy session (2005).

King (2013) compared the Jamar electronic dynamometer to the Jamar hydraulic dynamometer to determine reliability and validity. The study included 20 men and 20 women between the age of 20 and 50. The results indicated that the hydraulic dynamometer measured 11.27% higher for men and 8.88% higher for women than the electronic dynamometer. On average, the Jamar hydraulic dynamometer measured approximately 13 pounds higher for men and six pounds higher for women. The paired t-test p value was less than .0001 which demonstrates a statistically significant difference. Furthermore, the ICC values were .62 for men and .74 for women. ICCs above .75 were considered to indicate good reliability. The author concluded that clinicians should not interchange these two instruments with patients and that they should document which specific instrument was used with each patient (King, 2013).
What is lacking in the research and how will our study add to the literature?

Prior studies have reported mixed results in determining inter-instrument reliability among various dynamometers used to measure grip strength. Studies have tested a few digital dynamometers for other types of reliability and validity. Hinson, Woodard, and Gench (1990) conducted a study on the Jamar digital dynamometer model 2A. The sample included 50 right hand dominant 15 year old males. They found that the dynamometer showed high test-retest reliability (r=0.954 on the right hand and r= 0.946 on the left hand). Niebuhr, Marion, and Fike (1994) tested the computer connected Jamar dynamometer, model PC5030PT, for test-retest reliability. Participants included 33 healthy males and females age 20-41. The results showed good test-retest reliability (ICC between .80 and .90). However, neither study compared the instrument to the “gold standard” to establish concurrent validity or inter-instrument reliability. Existing normative data were established using the Jamar hydraulic dynamometer (Bohannon, Peolsson, & Massey-Westropp, 2006; Mathiowetz, Kashman, Volland, Weber, Dowe, & Rogers, 1985). Without established inter-instrument reliability of other types of dynamometers, clinicians are not justified in using the normative data to evaluate clients.

There have been no studies documenting the reliability and validity of the Baseline digital dynamometer. Fess (1986) stated, “it may not be assumed that an instrument is a reliable and valid assessment tool because it has a digital readout, or produces computerized notations” (p. 622). The purpose of our study is to examine the inter-instrument reliability and concurrent validity of the Baseline digital dynamometer to determine if it can be used interchangeably with the Jamar hydraulic dynamometer. Our study will answer the question of whether there is a difference in grip strength measurements between the Baseline digital dynamometer and the Jamar hydraulic dynamometer in healthy adults. Results of the study would have implications for
generalizing the published norms established for the Jamar hydraulic dynamometer to the Baseline digital dynamometer.

Method

Participant characteristics

Approval was given by The University of Toledo Institutional Review Board (IRB) to recruit human subjects for this study. We recruited participants who were between the ages of 18 and 50 years of age. This age range was used to eliminate variability in the grip strength measurements, as grip strength peaks between age 20-50 (Mathiowetz, Kashman, Volland, Weber, Dowe, & Rogers, 1985). Individuals less than 18 years of age can vary in their level of development. Individuals greater than 50 years of age may start experiencing physiological effects of aging (Horowitz, Tollin, & Cassidy, 1997). Both introduce a potential for increasing the variance in the grip strengths among these individuals, therefore the noted age range was used. Both males and females were included in the study. Participants were self-reported to be healthy and free from any disease, neuromuscular, or orthopedic injury that could affect their upper extremity strength. Cognition was not tested but it was observed that participants were able to understand directions and provide informed consent. It was also required that participants were English speaking.

Participants were recruited using word of mouth and personal approach, fliers, email invitation, and poster display. Participants were recruited from main campus of The University of Toledo. A total of 102 participants were recruited in accordance with the sample size of other studies (Mathiowetz, Vizenor, & Melander, 2000; Mathiowetz, 2002; Shechtman, Gestewitz, & Kimble, 2005).
Study design

This was a cross-sectional, descriptive study with a randomized crossover design and repeated measures. This manuscript is part of a larger study that compared the Baseline digital pinch gauge and the B&L hydraulic pinch gauge, in addition to the comparison of the Baseline digital dynamometer and the Jamar hydraulic dynamometer. Inter-instrument reliability and concurrent validity were examined for the Baseline digital dynamometer and Baseline digital pinch gauge. A computerized randomization program was used to randomize several factors to control for order effects (see figure 1). First, the order of pinch and grip was randomized, resulting in some participants being measured on grip first and some being measured on pinch first. The order of the three types of pinch measurements was also randomized. Last, the order of the type of instrument was randomized for both the dynamometers and pinch gauges. For example, some participants were measured on the digital instrument first and others were measured on the hydraulic instruments first. This controlled for type I error due to fatigue effects or other confounding variables. The chance of a type II error was also low with this design because each person was compared to him/herself, which eliminated dispersion due to individual differences (Nelson, 2006).

Measures/apparatus/instruments

We used the Jamar hydraulic dynamometer (model 5030J1) purchased from Lafayette Instruments (see figure 2) as the “gold standard” comparison instrument for this study. The Jamar hydraulic dynamometer measures a maximum of 200 pounds (90 kilograms) and has a peak-hold needle that retains the highest reading on the gauge until it is reset by the examiner. It requires an isometric contraction with no perceptible motion of the handles, regardless of the grip strength (Manual for the Jamar Hydraulic Hand Dynamometer, nd). The Baseline digital
dynamometer (Fabrication Enterprises, Inc., White Plains, NY) was the instrument that was compared to the Jamar hydraulic dynamometer (see figure 2). The Baseline digital dynamometer measures a maximum of 300 pounds (135 kilograms). This dynamometer features an LCD display for measurement read-out. The Baseline digital dynamometer features an electronic zero calibration system and a series of buttons that allow the user return to zero, store the last maximum reading, display the maximum reading from the stored memory, and the ability to exchange the measurement readout between pounds and kilograms (Fabrication Enterprises, Inc, 2011).

The Jamar hydraulic dynamometer has substantial psychometric evidence supporting its reliability and validity. Mathiowetz, Weber, Volland, and Kashman (1984) reported that the Jamar hydraulic dynamometer demonstrated good to excellent test-retest reliability ($r= 0.88$ to $0.93$), excellent interrater reliability ($r= 0.99$), and a $\pm 3\%$ accuracy. It has also been determined that occupational therapy students who are properly trained in the procedure for measuring grip strength demonstrate high inter-rater reliability with the Jamar hydraulic dynamometer (ICC=.996-.998) (Lindstrom-Hazel, Kratt, & Bix, 2009). There are currently no studies reporting on the psychometric properties of the Baseline digital dynamometer. Additionally, both dynamometers were adjusted to the second handle position during testing in order to optimize reliability and validity (Fess, 1992).

Prior to the start of data collection, the Baseline digital dynamometer was sent to Fabrication Enterprises, Inc. for calibration. The vendor certified that the calibration testing of the instrument was accomplished in accordance with the applicable specifications using test weights or equipment periodically calibrated with weights traceable to National Institutes of Standards and Technology. A new Jamar hydraulic dynamometer was purchased for this study;
therefore calibration of this instrument at the start was documented by the vendor. Both instruments were returned to the vendors halfway through data collection for re-calibration. All instruments were "appropriately calibrated", although any differences in instrument readouts were not relayed to us from the vendor.

**Procedure**

The participants were directed to sit at a designated table and chair upon entry into the testing area. The chair was standard height and did not have arm rests. The test administrator determined if the participant met the inclusion criteria by inquiring about his/her age and if he/she had any diseases, neuromuscular, or orthopedic injury that affects his/her upper extremity strength. If the participant met the inclusion criteria, he/she was then presented with an informed consent form. The test administrator explained the purpose of the study and risks that might be involved with participation. The participant was given the option to sign the consent form or to opt out of participation at this time, or any time during the testing session. After the participant signed the consent form, data collection began.

The study followed the American Society of Hand Therapists procedure for positioning, which requires a standard height chair without arm rests. The participant was “seated with [his/her] shoulder adducted and neutrally rotated, elbow flexed at 90 degrees, forearm in neutral position, and the wrist between 0 degrees and 30 degrees dorsiflexion and between 0 degrees and 15 degrees of ulnar deviation” (Fess, 1992). The following verbal instructions were given, “I want you to hold the handle like this and squeeze as hard as you can.” The examiner demonstrated and then gave the dynamometer to the participant. After the participant was positioned appropriately, the examiner said “Are you ready? Squeeze as hard as you can.” As the subject began to squeeze, the administrator said, “Harder!....Harder!....Relax”(Mathiowetz,
Weber, Volland, & Kashman, 1984, pp. 224). After the first trial score was recorded, the test was repeated with the same instructions for the second and third trial. The participant received a 15 second rest break between each trial. After completing all trials for the first dynamometer, the participant received a one minute rest break before the process was repeated for the second dynamometer. Three successive trials were completed on the right hand for both dynamometers, therefore a total of 6 grip strength trials were completed. The order of the dynamometers was determined by randomization however both were tested on the right hand. As this is part of a larger study, procedures for pinch gauge measurements are reported in a manuscript authored by Kailee Miller. As previously mentioned, pinch strength measurements were also part of the randomized process and may precede grip measurements. The entire data collection procedure lasted between 20 and 30 minutes.

**Data analysis**

Interval level data were collected. The mean of the three trials on each dynamometer was calculated. The mean of the Jamar hydraulic dynamometer and the mean of the Baseline digital dynamometer were compared for each participant. We used the intraclass correlation coefficient (ICC) and t-tests for paired measures to test for inter-instrument reliability and concurrent validity between the two instruments. We used a two-way mixed effects model to measure the ICC. The ICC ranges from 0.00 to 1.00 and measures the degree of agreement between the two instruments (Portney & Watkins, 2009). Values close to one indicate a high level of agreement (Portney & Watkins, 2009). Effect sizes were calculated with d=.20 considered small, d=.50 considered moderate, and d=.80 considered large (Cohen, 1988). An effect size is an estimate of the magnitude of difference between the two instruments (Portney & Watkins, 2009). We also completed a Bland-Altman analysis on the data. This type of analysis compares two
measurement methods by computing a bias (the average of the differences). Values close to zero indicate that the two methods are producing similar results (Portney & Watkins, 2009). Significance level was set at an alpha value of less than 0.05.

**Results**

**Participants**

The ages of the 102 study participants ranged from 18-50 years. The mean age of participants was 25.51 (SD=7.25) and the median age was 23.00. More than half (57.8%) of the participants were within the range of 21-24 years old. Of the participants, 54 (52.9 percent) were male and 48 (47.1 percent) were female. The participants consisted of 88 individuals who were right hand dominant (86.3 percent), 11 participants who were left hand dominant (10.8 percent), and 3 participants who used both hands equally (2.9 percent). Among the study sample, 92.2% of the participants were Caucasian (n=90), 5.9% of participants were African American (n=6), 2.9% were of Hispanic descent (n=3), and 1.0% were of Asian descent (n=1).

**Descriptive Statistics**

No data were missing and analyses indicated that there were no significant order effects for the paired data. T-tests for paired data indicated that there was a significant difference between the measurements of the two instruments ($t=2.60; p=.011$). The mean measurement from the Jamar hydraulic dynamometer was 1.80 lbs. greater than the mean measured by the Baseline digital dynamometer (91.25 lbs. and 89.45 lbs. respectively). The grip strength measurements for the Jamar hydraulic dynamometer ranged from 51.67 lbs. to 176.33 lbs. and from 47.00 lbs. to 174.00 lbs. on the Baseline digital dynamometer. There is a greater range of values measured by the Baseline digital dynamometer which indicates a greater amount of variability in the measurements as compared to the Jamar hydraulic dynamometer (see table 1).
The standard deviation for both mean values is similar which adds strength to our study because it means that the variability is relatively even across the data (see table 1). We can be 95% confident that the true mean difference between the instruments lies somewhere between .425 and 3.163 lbs. ($t=2.60$). The 95% confidence interval does not contain zero, which indicates instrument bias. The ICC value was .982, which indicates that the data from both dynamometers were highly correlated. Finally, the effect size was very small ($d=0.07$).

The Bland-Altman analysis indicated a bias (difference between the means) of 1.79. This indicates that the two dynamometers are producing different results because with this analysis values close to zero indicate no difference between the differences in the means. In 95% of participants the average difference lies between -11.87 and +15.46 (see figure 3). This means that the grip strength measurements can vary up to 27.32 lbs. between the instruments. Although the Bland-Altman plot (see figure 3) appears to be unbiased, data from the Bland-Altman analysis indicated that the Jamar hydraulic dynamometer measured higher in 67 participants whereas the Baseline digital dynamometer measured higher in 35 participants. Therefore, on average the Jamar produces higher measurements than the Baseline digital dynamometer.

Discussion

This study investigated the inter-instrument reliability and concurrent validity of the Baseline digital dynamometer. The purpose of the study was to determine whether there is a difference in grip strength measurements between the Baseline digital dynamometer and the Jamar hydraulic dynamometer in healthy adults. Researchers have described the importance of using valid and reliable instruments in clinical practice (Fess, 1986). However, until now, no studies have tested the Baseline digital instruments for reliability and validity. Our results indicate that there is a statistically significant difference in the grip strength measurements.
between the Baseline digital dynamometer and Jamar hydraulic dynamometer. On average, the Jamar hydraulic dynamometer measures higher than the Baseline digital dynamometer.

The high correlation between the Jamar hydraulic dynamometer and the Baseline digital dynamometer indicates that the two instruments are measuring the same construct of grip strength. Despite the excellent association between the two instruments, they do not show strong concurrent validity. Several studies have found similar results for various instruments including the Baseline hydraulic dynamometer, the Biometrics E-LINK, and the DynEx digital dynamometer (Mathiowetz, Vizenor, & Melander, 2000; Allen & Barnett, 2011; Shechtman, Gestewitz, & Kimble, 2005). Therefore, it is possible to find statistically significant differences between the scores of the two instruments if one instrument demonstrates consistently higher scores and yet still obtain a high intraclass correlation coefficient (inter-instrument reliability) between the two instruments (Shechtman, Gestewitz, & Kimble, 2005).

Although the results of this study indicated that there were statistically significant differences, we also found that the effect size was very small. This indicates that while there may be differences between the two instruments, perhaps the differences are small between the two instruments. This is expected in a study of healthy adults where the average grip strength measured was around 90 lbs. and the maximum grip strength measured as high as 175 lbs. (see table 1). In grip strength measurements of this magnitude a difference of 1.80 lbs. is inconsequential and would not have an impact on a person being able to perform a given task. However, we must consider that typical occupational therapy clients being evaluated on grip strength are likely to have values well below the norm for their age and gender due to disease and/or injury. A 1.80 lbs. difference will be more significant among lower grip strength values and this will include the vast majority of clients with hand injuries.
Comparison of the grip strength measurements obtained with the two instruments showed that the bias between the two dynamometers was fairly small in terms of hand strength, however, the large span of the limits of agreement indicates that the instruments cannot be exchanged for ongoing measurements of the same patient, nor can values be translated between the dynamometers (Massey-Westropp, Ahern, & Hearn, 2004). Massey-Westropp, Ahern, and Hearn (2004) revealed an average difference of 4.95 lbs. and a range of 48.33 lbs. between their limits of agreement between the electronic Grippit and the Jamar hydraulic dynamometer. Svens and Lee (2005) found an average difference of 6.06 lbs. and the limits of agreement ranged a total of 22.49 lbs. Although the average difference in our study was smaller than both of these studies, our limits of agreement were larger than the results of Svens and Lee (2005) who concluded that differences of this magnitude were unacceptable for clinical practice. The spread around the zero point (see figure 3) helps us decide that the error is not acceptable if we substitute the Baseline digital dynamometer for the Jamar hydraulic dynamometer and vice versa.

Our results differ from a study by Mathiowetz (2002) that compared the Roylan and Jamar hydraulic dynamometers. Their study found that the two instruments had acceptable concurrent validity and excellent inter-instrument reliability. Therefore, the researcher concluded that the Jamar and Rolyan hydraulic dynamometers measure equivalently for practical purposes and can be used interchangeably as well as with normative data established for the Jamar dynamometer (Mathiowetz, 2002). However, the author does caution that the ability to interchange grip dynamometers should never be assumed. “[The Mathiowetz] study used one Jamar and one Rolyan dynamometer. It cannot be assumed that all Jamar and all Rolyan dynamometers measure equivalently unless their concurrent validity is acceptable” (Mathiowetz, 2002, p. 208). Similarly, Flood-Joy and Mathiowetz (1987) have provided evidence that different
versions of the Jamar hydraulic dynamometer do not always measure equivalently, therefore it cannot be assumed that all Baseline digital dynamometers read equivalently. Thus, occupational therapists are encouraged to always use the same dynamometer when pre- and post-testing clients (Flood-Joy & Mathiowetz, 1987; Shechtman, Gestewitz, & Kimble, 2005; Allen & Barnett, 2011).

Implications for Occupational Therapy

An average difference of 1.80 lbs. between dynamometers may not make a difference in terms of being able to carry out a specific grip strength task in everyday life, such as opening a container in healthy adults. However, for an individual with significant weakness, 1.80 lbs. could have a major impact on carrying out certain tasks as Rice, Leonard, & Carter found that it required as little as 2.31 lbs. to use a trigger pump spray bottle (1998). In a clinic setting there is an emphasis on being able to document change or improvements in our clients. In this scenario, when a therapist is documenting a client’s change in grip strength over a period of time, a difference in 1.80 lbs. has the potential to introduce a meaningful amount of variability in the clinical documentation. It should also be taken into consideration that the values can potentially vary up to 27.32 lbs. Our results indicate that therapists should not use these two instruments interchangeably with the same client. It is not valid to compare numbers obtained with the Baseline digital dynamometer to the norms, which were established with the Jamar hydraulic dynamometer as these two instruments do not consistently measure the same.

Strengths and Limitations

This study had an equal representation of males and females, however, we did not stratify the data based on gender as several other studies did (Mathiowetz, Vizenor, & Melander, 2000; Massy-Westropp, Ahern, & Hearn, 2004). Therefore, we do not know if the differences occurred
equally among male and female participants. Over half of the participants in our study were college age individuals. Therefore, our study does not reflect the diverse age representation of the general adult population. Our study population lacks ethnic diversity so we should take caution with generalizing the results to the overall population. The Jamar dynamometer peak hold needle started malfunctioning halfway through the research project and needed to be sent to the company that it was purchased from to be repaired and re-calibrated. Finally, we did not do statistical analyses during the calibration check half-way through the study. We do not know if there was a significant amount of measurement error (intra-instrument reliability) throughout the study.

**Future Research**

Future studies testing the reliability and validity of grip and pinch strength instruments should have an equal distribution of ages within the range of 18-50 years. Investigations should be conducted to establish normative data using the Baseline digital dynamometer. This would give therapists a set of reference values that they could compare client scores to in order to indicate severity of impairment in a valid and reliable way. Other aspects of grip strength instrumentation that would be interesting to test include, comparisons of the Baseline digital dynamometer with the current gold standard in terms of ease to use, length of testing time, and accuracy and precision (intra-instrument reliability). During our study, participants commented on the differences in how each dynamometer felt in their hands during testing. Future research could examine participant perceptions of which instrument is more preferable as well as anthropometric impact on grip strength. Furthermore, some participants felt motivated to grip harder than their previous trial, so it would be interesting to investigate the impact of visual feedback on motivation during grip strength testing.
Conclusion

Using assessment instruments that are reliable and valid is important when measuring and documenting grip strength in occupational therapy. This study indicates that therapists can use the Baseline digital dynamometer and know that it is reliably measuring the construct of grip strength. However, because of the significant differences found in the grip strength readings between the two instruments, we recommend that the two dynamometers should not be interchanged and that the same dynamometer should be used to measure grip strength of a client throughout his/her treatment program. Furthermore, therapists should not compare scores obtained with the Baseline digital dynamometer to the normative standards that are currently available.

Acknowledgements

There are several individuals who have contributed in substantial ways during the process of completing this scholarly project. First, I would like to sincerely express my appreciation to my research advisor, Julie Jepsen Thomas, Ph.D., OTR/L, FAOTA for sharing her expertise, providing support, giving insight, and offering encouragement throughout the entire process. Kailee Miller, for working alongside me as my research partner and dedicating time to research meetings, participant recruitment, data collection, and presenting at research conferences. Martin Rice, Ph.D, OTR/L for assisting with data analysis and patiently explaining the data from the statistical analysis. The contributions of those mentioned were greatly valued. They have all positively influenced the overall quality and implications that this project contributes to the profession of occupational therapy.
References


Manual for the Jamar Hydraulic Hand Dynamometer, model 5030J1, Sammons Preston Roylan, 4 Sammons Court, Bolingbrook, IL, 60440


### Table 1

*Descriptive Data for Dynamometers*

<table>
<thead>
<tr>
<th></th>
<th>M (SD)</th>
<th>SEM</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamar</td>
<td>91.25 (26.22)</td>
<td>2.60</td>
<td>51.67</td>
<td>176.33</td>
<td>124.67</td>
</tr>
<tr>
<td>Baseline</td>
<td>89.45 (26.95)</td>
<td>2.67</td>
<td>47.00</td>
<td>174.00</td>
<td>127.00</td>
</tr>
</tbody>
</table>

*Note. n=102, all figures are displayed in pounds*
Figure 1. Randomization program for grip and pinch order.
Figure 2. Jamar® hydraulic dynamometer (left) and Baseline® digital dynamometer (right)
Figure 3. Bland-Altman Plot for Gross Grasp measurements using the Jamar and Baseline dynamometers. Differences in measurements between the two dynamometers are plotted against the averages for 102 participants. Bias is indicated by the middle line (1.79). Most participants fell between the limits of agreement (outer lines) with more positive data points (67) than negative data points (35). Outer lines indicate 2 standard deviations above and below the mean.