Interrater Reliability of the MED-AUDIT (Medical Equipment Device – Accessibility and Universal Design Information Tool)

Rochelle J. Mendonca[1], Maysam M. Ardehali[2], Sayeda Farzana Akhtar[3], & Roger O. Smith[2]

[1]Columbia University, New York, NY, 10032
[2]Rehabilitation Research Design and Disability (R_2D_2) Center, University of Wisconsin-Milwaukee Milwaukee, WI, 53211
[3]Marquette University, Milwaukee, WI, 53211

ABSTRACT

The MED-AUDIT (Medical Equipment Device – Accessibility and Universal Design Information Tool) prototype demonstrates the potentials of this measurement approach to assess the accessibility of medical devices. This study investigated the inter-rater reliability of the MED-AUDIT. It was hypothesized that MED-AUDIT would demonstrate good to excellent interrater reliability across five different models of blood pressure monitors, one manual inflate arm monitor, two automatic arm monitors, and two automatic wrist monitors when scored on the MED-AUDIT by 3 different raters. Intraclass correlation coefficients were calculated across the three raters for 1) the raw scores of the top two parent levels of questions in the taxonomy and 2) summary accessibility scores for 13 impairments. Results showed the MED-AUDIT demonstrated good to excellent interrater reliability for all devices, except that it was moderate for raw parent questions for the automatic arm monitor and was poor for the 13 impairment categories for the manual blood pressure monitor. This study demonstrates that the MED-AUDIT evaluation methodology is feasible and reliable, however requires further testing.

INTRODUCTION

In the past few decades there has been a dramatic increase in the survival rate of individuals with severe disabilities and an increased demand for medical devices to be used in patient's homes, with the patient as the end user [1-5]. However, tremendous disparities in health care access prevent these populations from receiving treatment in a timely manner, primarily due to inaccessible medical technologies [6-8]. For example, to monitor their blood pressure, individuals must be able use their upper extremities to get the cuff on themselves and to see or hear the display; or individuals must be able to stand to get a mammogram; or must have adequate strength to climb and balance on an examination table. This translates into hundreds of millions of dollars spent by clinical professionals, caregivers, and lay people to purchase medical devices that may not work for a large majority of the patients they are intended for, thereby creating a large disparity in healthcare access [9-11].

Three factors exacerbate the problem and highlight the push towards accessible medical devices: (1) the rapid aging of America, (2) increased survival rate for people with disabilities (PWD), and (3) increase in use of home health equipment used and purchased directly by patients [12-14]. However, there is no assessment that measures the accessibility of medical devices. The MED-AUDIT is being developed to meet this need. The MED-AUDIT is a software-based assessment that performs a task analysis, assesses design features, addresses information for specific impairments, and calculates an integrated accessibility score for medical devices. The questions on task analysis and design features comprise the taxonomy of the MED-AUDIT. It includes two background matrices that represent relationships between impairments and device features and device features and tasks which create a database map to produce accessibility scores for thirteen different impairments.

Past studies have reported on aspects of the development, and satisfactory usability, feasibility, and preliminary psychometric properties of the MED-AUDIT [15]. It has already been established that MED-AUDIT distinguishes between accessibility levels for different devices [16], however the inter-rater reliability of the MED-AUDIT across real devices across multiple raters has not been investigated. Therefore, the research question explored in this paper is, "Does the MED-AUDIT taxonomy demonstrate inter-rater reliability across five different models of blood pressure monitors across three raters"?

METHODS

This study was designed to determine the inter-rater reliability of the MED-AUDIT taxonomy across five different models of blood pressure monitors (BPM), when scored individually by 3 trained raters (See Figure 1). The five BPM rated in this study consisted of a manual inflate blood pressure monitor, two automatic arm monitor, and two automatic wrist monitors. The BPM were chosen to represent the range of blood pressure monitoring devices available as well as to represent different levels of accessibility.

All five BPM were scored on the same version of the MED-AUDIT, which includes about 1150 distinct questions arranged in a hierarchal outline broken down into five or six levels for the taxonomy. The outline structure provides the branching options from level to level. The branching enables irrelevant parts of the taxonomy for a certain device to be bypassed, making questions targeted and requiring less expertise (See Figure 2). The taxonomy includes 193 questions related to tasks required to use the medical device and are scored on a three-point scale (0-does not require, 1-somewhat requires, and 2-requires). It also includes 957 questions related to features included in the device and are scored on a 3-point scale (0-does not include, 1-somewhat includes, and 2-includes).

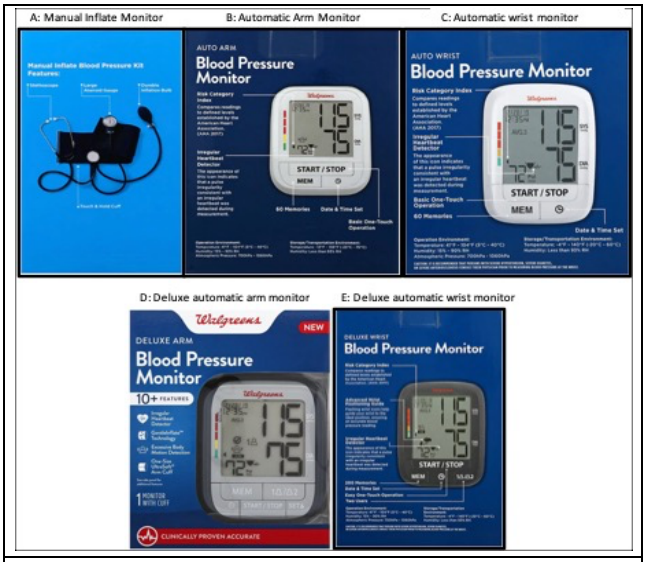

Figure 1: Four models of Blood Pressure Monitors

Figure 2: Sample of MED-AUDIT scoring taxonomy

Three raters unpackaged and interacted with each device prior to scoring it. Each device was scored independently before moving onto the next device. On completion of scoring, the data from the software was exported to an Excel sheet which provided a raw score and a percentage score for each question. We also obtained summary accessibility scores across 13 impairment categories: hard of hearing, deaf, low vision, blind, expressive communication, comprehensive disorders, other cognitive disorders, behavioral impairment, sensitivity impairment, lower limb impairment, upper limb impairment, head/neck/trunk impairment, and systemic body impairment. Intraclass correlation coefficients were calculated across the three raters for each device for 1) the raw scores of the top two parent level questions in the taxonomy and 2) summary accessibility scores for the 13 impairments. ICC is interpreted as >0.90 = excellent reliability, 0.75-0.9 = good reliability, 0.5-0.75 = moderate reliability, and <0.5 = poor reliability [17]. In addition, Gwet's AC1 was used to assess agreement on the smallest child level questions for a more robust evaluation of interrater reliability.

RESULTS

The raters took between 45 minutes to an hour to score each of the five devices, including unpackaging the device, reviewing instructions, interacting with the device, and rating the MED-AUDIT for the device. The rater reported no difficulties using the MED-AUDIT. Table 1 highlights the results for intraclass correlation coefficients for the raw scores across the top two parent levels of questions of the MED-AUDIT taxonomy for all five blood pressure monitors. The scores for the Automatic wrist (ICC = 0.957) and Deluxe arm monitor (ICC = 0.922) were excellent, for the Manual (ICC = 0.844) and Deluxe wrist monitor (ICC = 0.957) were good, and for the Automatic arm monitor was moderate (ICC = 0.565).

Table 1: Intraclass Correlation Coefficients Across Tasks and Device Features

Manual Blood Pressure Monitor	0.844
Automatic Arm Blood Pressure Monitor	0.565
Automatic Wrist Blood Pressure Monitor	0.957
Deluxe Arm Blood Pressure Monitor	0.922
Deluxe Wrist Blood Pressure Monitor	0.825

Table 2: Intraclass Correlation Coefficients Across 13 Impairment Categories

Manual Blood Pressure Monitor	0.311
Automatic Arm Blood Pressure Monitor	0.928
Automatic Wrist Blood Pressure Monitor	0.978
Deluxe Arm Blood Pressure Monitor	0.949
Deluxe Wrist Blood Pressure Monitor	0.966

Table 2 demonstrates the results for the ICC across the three raters across all 13 impairments for the five blood pressure monitors. The ICC's were excellent for the Automatic arm (ICC = 0.928), the Automatic wrist (ICC = 0.978), the Deluxe arm (ICC = 0.949), and the Deluxe wrist (ICC = 0.966) monitors but was poor for the Manual monitor (ICC = 0.311). Figure 3 presents line graphs of the three raters scores for the five devices across the 13 impairments. The graphs highlight that ratings for the manual blood pressure monitor were varied across the 3 raters compared to the other four monitors.

Table 3 shows Gwet's AC1 agreement coefficients at the level of the lowest level of child questions. The responses to these questions are directly provided by the raters. AC1 coefficients were good for Automatic Arm Blood Pressure Monitor (AC1 = 0.816), Automatic Wrist Blood Pressure Monitor (AC1 = 0.848), Deluxe Arm Blood Pressure Monitor (AC1 = 0.822), and Deluxe Wrist Blood Pressure Monitor (AC1 = 0.860). This coefficient shows poor reliability for the Manual Blood Pressure Monitor (AC1 = 0.254).

Figure 3: Line Graphs of Rater Scores Across 13 Impairment Categories for Five BPMs

DISCUSSION

Results from the ICC and AC1 analyses across three raters show that the coefficients were good to excellent for all devices, except that it was moderate for the tasks and device features for the Automatic arm monitor and poor across accessibility scores for the 13 impairment categories as well as for child level analyses for the manual blood pressure monitor. In looking at the graphs and scores across raters, although a majority of analysis demonstrate good to excellent reliability, we would like to establish excellent reliability irrespective of the complexity of the device. The results highlight a need to conduct a further detailed analysis of the data at the question level to determine which areas led to differences between rater scores. We will also evaluate the question descriptions to determine if they need additional detail to clarify the concepts. Last, we will investigate and evaluate the matrices and algorithms that generate accessibility scores to determine if they need revision.

Table 3: Gwet's AC1 Agreement Coefficients Across Smallest Child Level Questions	
Manual Blood Pressure Monitor	0.254
Automatic Arm Blood Pressure Monitor	0.816
Automatic Wrist Blood Pressure Monitor	0.848
Deluxe Arm Blood Pressure Monitor	0.822
Deluxe Wrist Blood Pressure Monitor	0.860

This study suggests some future steps for continuing to establish the reliability and validity of the MED-AUDIT: 1) test different types of medical devices that cover the spectrum of diagnostic, monitoring, positioning, and maintenance devices, and 2) re-evaluate the questions and complete a detailed analysis of the appropriateness and accuracy of questions towards measuring medical device accessibility.

Establishing the reliability and validity of the MED-AUDIT scores will provide insights into how to provide this information to designers as well as consumers of medical devices and may have labeling implications for medical devices. Providing consumers with accessibility information will improve medical device design and healthcare for consumers with disabilities.

REFERENCES

[1] Gans, B. M., Mann, N. R., & Becker, B. E. (1993). Delivery of primary care to the physically challenged. *Archives of Physical Medicine and Rehabilitation, 74*(supplement), s15-s19.

[2] Kaye, H. S., Laplante, M. P., Carlson, D., & Wegner, B. L. (1996). Trends in disability rates in the United States, 1970-1994. *Disability Statistics Abstract, 17,*1-6.

[3] Wilcox, S. B. (2003). Applying the principles of universal design to medical devices. *MDDIOnline.* Retrieved October 11, 2004, from https://www.mddionline.com/news/applying-universal-design-medical-devices.

[4] Story, M.F., Winters, J.M., Kailes, J.I., Premo, B., Winters, J.M. (2003). Understanding Barriers to Healthcare Caused by Inaccessible Medical Instrumentation. *Proc. RESNA 2003 Annual Conf*, June18, 2003.

[5] Winters, J.M., Story, M.F., Kailes, J.I., Premo, B., Danturthi, S., Winters, J. (2004). Accessibility of Medical Instrumentation for Persons with Disabilities: A National Survey. *Midwest Nursing Reseach Society* , St. Louis, MO, February 27-March 1, 2004.

[6] Cheng, E., Myers, L., Wolf, S., Shatin, D., Cui, X., Ellison, G., Belin, T., & Vickrey, B. (2001). Mobility impairments and use of preventive services in women with multiple sclerosis: observational studies. *British Medical Journal, 323*(7319), 968–969. https://doi.org/10.1136/bmj.323.7319.968

[7] Schopp, L., Sanford, T., Hagglund, K., Gay, J., & Coatney, M. (2002). Removing service barriers for women with physical disabilities: Promoting accessibility in the gynecologic care setting. *Journal of Midwifery & Women's Health, 47*(2), 74–79. https://doi.org/10.1016/S1526-9523(02)00216-7

[8] Veltman, A., Stewart, D., Tardif, G., & Branigan, M. (2001). Perceptions of primary healthcare services among people with physical disabilities. Part 1: access issues. *MedGenMed : Medscape General Medicine*, *3*(2), 18

[9] Grabois, E., Nosek, M. A., & Rossi, D. (1999). Accessibility of primary care physicians' offices for people with disabilities. *Archives of Family Medicine*, *8*(1), 44–51. https://doi.org/10.1001/archfami.8.1.44

[10] North Carolina Office on Disability and Health (NCODH). (2007). *Removing barriers to health care: A guide for health professionals*.

[11] Markwalder, A. (2005). *Disability rights advocates. In A call to action: A guide for managed care plans serving Californians with disabilities*.

[12] Gans, B. M., Mann, N. R., & Becker, B. E. (1993). Delivery of primary care to the physically challenged. *Archives of Physical Medicine and Rehabilitation*, *74*(Suppl), s15–s19.

[13] Kraus, L., Lauer, E., Coleman, R., & Houtenville, A. (n.d.). *2017 Disability Statistics Annual Report*.

[14] R Sade, R. M. (2012). The graying of America: Challenges and Controversies. *The Journal of Law, Medicine, & Ethics*, *40*(1), 6–9. https://doi.org/https://doi.org/10.1111/j.1748-720X.2012.00639.x

[15] Mendonca, R., & Smith, R. O. (2007). Validity analysis: MED-AUDIT (Medical Equipment Device-Accessibility and Universal Design Information Tool). *RESNA 30th International Conference on Technology and Disability: Research, Design, Practice and Policy*.

[16] Mendonca, R. J., O'Donnell, L. M., & Smith, R. O. (2021). Feasibility analysis: MED-AUDIT (Medical Equipment Device-Accessibility and Universal Design Information Tool). *RESNA 2021 International Conference on Technology and Disability: Research, Design, Practice and Policy*.

[17] Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012