

Assessing the Usability of Complex Psychosocial Interventions: The Intervention Usability Scale

Aaron R Lyon (✉ lyona@uw.edu)

University of Washington <https://orcid.org/0000-0003-3657-5060>

Michael D. Pullmann

University of Washington

Jedediah Jacobson

University of Washington

Katie Osterhage

University of Washington

Morhaf Al Achkar

University of Washington

Brenna N. Renn

University of Washington

Sean A. Munson

University of Washington

Patricia A. Areán

University of Washington

Short report

Keywords: Complex health interventions, Psychosocial interventions, Mental health, Usability, Human-centered design, Primary care

Posted Date: November 9th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-20758/v3>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on February 8th, 2021. See the published version at <https://doi.org/10.1177/2633489520987828>.

Abstract

Background. Usability – the extent to which an intervention can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction – is a key determinant of implementation success. However, few instruments have been developed to measure the design quality of complex health interventions (i.e., those with several interacting components). This study evaluated the structural validity of the Intervention Usability Scale (IUS), an adapted version of the well-established System Usability Scale (SUS) for digital technologies, to measure the usability of a leading complex psychosocial intervention, Motivational Interviewing (MI), for behavioral health service delivery in primary care. Prior SUS studies have found both one- and two-factor solutions, both of which were examined in the current study of the IUS.

Method. A survey administered to 136 medical professionals from 11 primary care sites collected demographic information and IUS ratings for MI, the evidence-based psychosocial intervention that primary care providers reported using most often for behavioral health service delivery. Factor analyses replicated procedures used in prior research on the SUS.

Results. Analyses indicated that a two-factor solution (with “usable” and “learnable” subscales) best fit the data, accounting for 54.1% of the variance. Inter-item reliabilities for the total score, usable subscale, and learnable subscale were $\alpha = .83$, $\alpha = .84$, and $\alpha = .67$, respectively.

Conclusions. This study provides evidence for a two-factor IUS structure consistent with some prior research, as well as acceptable reliability. Implications for implementation research evaluating the usability of complex health interventions are discussed, including the potential for future comparisons across multiple interventions and provider types, as well as the use of the IUS to evaluate the relationship between usability and implementation outcomes such as feasibility.

Background

Although complex health interventions (CHIs)—those with several interacting components—are common in contemporary healthcare worldwide (Campbell et al., 2007; Craig et al., 2013), clinicians often face considerable challenges implementing these interventions, owing to their design complexity (Greenhalgh et al., 2004; Perez Jolles et al., 2019). For this reason, intervention usability—the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction (ISO, 1998)—has been identified as a key “upstream” determinant of implementation (Lyon & Bruns, 2019). This is particularly true in mental and behavioral health, where most effective practices are complex evidence-based psychosocial interventions (EBPIs) (IOM, 2015).

In contrast to standard perceptual implementation outcomes such as acceptability, appropriateness, and feasibility, usability is largely a characteristic of the *intervention*. Lyon, Brewer, and Areán (in press) recently articulated a spectrum of contextual dependence for these four overlapping but distinct constructs, spanning from usability (the most contextually independent) to feasibility (the most contextually dependent). As an intra-intervention implementation determinant, usability is theorized to predict both perceptual (e.g., feasibility) and behavioral (e.g., adoption, fidelity, sustainment) implementation outcomes. However, despite the existence of over a hundred instruments available to assess standard implementation outcomes (Lewis et al., 2015a), no measures exist to evaluate CHI usability or other aspects of design quality (Lewis et al. 2015b). This limited instrumentation has inhibited empirical study of the construct, including examination of the relationships between usability and perceptual implementation outcomes.

Frequently, EBPIs are delivered in integrated or non-specialty care settings (e.g., primary care, schools) that differ markedly from the contexts where they were originally developed. For instance, mismatches between EBPI design and the constraints of primary care make intervention fidelity and patient outcomes difficult to sustain (Alexopoulos et al., 2016; Areán et al., 2008). Despite being critical to implementation, aspects of intervention design quality, such as usability, have been insufficiently assessed in primary care and other non-specialty contexts (Lyon et al., 2019).

Human-Centered Design (HCD) and Usability

Human-centered design (HCD) reflects an approach and set of methods for creating and refining systems so that they are usable and useful for their stakeholders. HCD is also closely related to the fields of human-computer interaction, user experience, and human factors (ISO, 2019; Norman & Draper, 1986; Rubin & Chisnell, 2008). Relevant to our research aims, each of these fields includes a focus on usability and provides methods to evaluate and improve the extent to which people can reliably use a system to achieve their goals, without error, safely, and with an enjoyable experience (Dumas, Dumas, & Redish, 1999; Neilsen, 1994). Usability is a key aspect in the design and refinement of CHIs (Burchett, Blanchard, Kneale, & Thomas, 2018; Harte et al., 2017; Horsky et al., 2012), and EBPIs in particular (Lyon et al., 2019; Lyon, Koerner, & Chung, 2020).

System Usability Scale (SUS). Originally designed for assessing digital systems, the System Usability Scale (SUS) is among the most widely applied usability instruments (Sauro & Lewis, 2009), and has been found to be significantly related to task success in both laboratory and field studies (Kortum & Peres, 2014). Some applications of the SUS have even extended beyond digital products to technologies such as automatic teller machines and microwave ovens (Kortum & Bangor, 2013). Brooke (1986) initially developed the 10-item scale as a brief and reliable instrument to allow comparisons among products and across versions of a product. Bangor et al. (2008) evaluated 2324 SUS questionnaires and found high internal reliability ($\alpha=.91$) and good sensitivity and concurrent validity. Although the instrument was originally considered unidimensional, Sauro and Lewis (2009) evaluated the SUS factor structure using 2648 questionnaires assessing products such as computer hardware, software, and websites. They identified an eight-item “Usable” subscale and a two-item “Learnable” subscale, a finding which provoked several replication studies (Kortum & Sorber, 2015; Lewis et al., 2013; Lewis, Brown, et al., 2015; Lewis, Utesch, et al., 2015; Sauro & Lewis, 2011), including one analyzing over 9,000 surveys (Lewis & Sauro, 2017). These studies did not replicate the initial findings. Although most studies found a two-factor solution, factors were defined by whether the item was positively worded or negatively worded, which Lewis and Sauro (2017) interpreted as practically unidimensional. The current study replicates these factor analyses, but using a different population (healthcare professionals instead of consumers) and different target (EBPIs instead of computer hardware/software).

Current Aims

In light of the criticality of CHI usability, and the paucity of psychometrically-sound instruments for its evaluation, we revised the SUS to evaluate EBPI usability. The current study assessed the revised instrument’s structural validity using confirmatory factor analytic methods. This study was carried out with primary care providers given the importance of primary care for the delivery of contemporary mental health services worldwide (Centers for Medicare & Medicaid Services, 2018; WHO, 2016) and the likelihood that many traditional EBPIs may demonstrate problematic usability in that novel context. The IUS was evaluated using primary care providers’ ratings of Motivational Interviewing, a client-centered and directive approach designed to help service recipients resolve ambivalence and build motivation to change (Miller & Rollnick, 2012). MI has a strong evidence base, including in primary care (Hettema, Steele, & Miller, 2005; VanBuskirk & Wetherell, 2014), but its usability has never been assessed.

Method

Participants

Table 1 provides sample demographics. The study sample consisted of 136 medical providers who selected Motivational Interviewing (MI) as the intervention they delivered most often (see Procedures). A slight majority (56.6%) of participants self-reported their gender as female; most participants were White (86.0%); and the median age was between 30-39.

Procedures

The data in the current study were collected as a component of a larger survey completed by the WWAMI (Washington, Wyoming, Alaska, Montana, Idaho) region Practice and Research Network (WPRN) – a group of primary care clinics committed to conducting practical research – in January 2019. At the time of the survey, WPRN clinics were spread across 25 diverse parent organizations in both urban and rural communities. The survey was originally conducted for quality improvement purposes. This study's analyses were approved by the authors' Institutional Review Board. The full survey consisted of questions in four areas: (1) clinician role/demographics, (2) perspectives on behavioral health in primary care, (3) anxiety treatment, and (4) provider burnout/resilience. Survey completion took approximately ten minutes.

All WPRN member organizations ($n=25$) were invited to participate. Consistent with standard practices in the WPRN, staff and faculty held webinars with WPRN clinics and a preview of the full survey was distributed to build engagement and secure leadership approval. Eleven clinic sites chose to participate. The survey was distributed via the REDCap (Harris et al., 2009) digital system by clinic champions to 491 providers representing three categories of primary care clinicians: medical providers (MD/DO, PA, NP), behavioral health providers (social workers, licensed addiction counselors, psychologists, marriage and family therapists), and clinical pharmacists. Participation was voluntary and no incentives were provided. The overall survey response rate was 56.8% ($n=279$). There were 205 (73.4%) participants qualified to answer the IUS questions, because they reported using at least one of six identified behavioral health interventions (see Measures). MI was chosen by the overwhelming majority of respondents ($n = 172$, 83.9%), and of those, most were medical providers ($n = 136$, 79.1%). Therefore, we selected only medical providers who chose MI as their primary behavioral health intervention to include in this study.

Measures

Survey of commonly used behavioral health interventions. As a component of the larger survey, respondents were asked to select which intervention they provided most frequently from a list of evidence-based behavioral health interventions known to be commonly used in the WPRN (including cognitive behavioral therapy, behavioral activation, and MI, among others). That intervention served as the referent for the IUS.

Intervention Usability Scale (IUS). The 10-item SUS (Brooke, 1986) was adapted to create the IUS (Lyon, 2016; Lyon et al., 2020). The term "system" was replaced with "intervention" in each item, but no other modifications were made to maintain consistency with the SUS operationalization of usability. Items are rated on a Likert scale from 0 (strongly disagree) to 4 (strongly agree), with half of the items reverse-scored. The total score was calculated by multiplying the sum of these scores by 2.5 (possible range: 0-100).

Analyses

To allow for direct comparability of studies, we replicated prior procedures. To replicate Sauro and Lewis (2009), we used principal axis factoring (a.k.a., common factor analysis) and, to replicate Lewis and Sauro (2017), we used principal components and unweighted least squares factor analysis to examine the factor structure of the IUS and evaluate possible subscales. All analyses used varimax rotation. We then assessed scale correlations and internal consistency for each subscale and the total score. We compared IUS descriptive statistics with prior research on the SUS.

Results

Item Descriptives and Correlations

Scores on IUS items included the full range (0-4) for all items besides item 4 (range: 0-3), item 5 (range: 1-4), and item 9 (range: 1-4). The mean scores for items ranged from 0.88 – 3.04; median scores were 3 for all regularly scored items and 1 for all reverse scored items. See Table 2 for more detail on individual item descriptive statistics. Inter-item correlation absolute values (half of the IUS items are reverse scored) ranged from 0.05 to 0.64. Five pairs of items were not significantly correlated: items 1 and 6, items 1 and 7, items 1 and 10, items 6 and 9, and items 6 and 10. See Table 3 for full item correlation results.

IUS Factor Structure

Examining the point of inflection on a scree plot indicated that a two-factor solution best fit the data. Factor 1 had an eigenvalue of 4.31, therefore accounting for 43.1% of the variance, and Factor 2 had an eigenvalue of 1.1, accounting for 11.0% of the variance, for a total of 54.1% of variance accounted for in the two-factor solution. For all three approaches to analysis, items 1, 2, 3, 5, 6, 7, 8, and 9 loaded onto the first factor, whereas items 4 and 10 loaded onto the second. See Table 4 for the rotated component matrix. The item-factor alignment was nearly identical to that of Sauro and Lewis (2009), which found a two-factor structure, but inconsistent with studies that followed (Lewis & Sauro, 2017). Therefore, we named these factors the same as in the 2009 study; “Usable” and “Learnable.” To place the Usable and Learnable scores on a comparable 0 to 100 scale as the Overall IUS score, we multiplied their summed score contributions by 3.125 and 12.50, respectively.

Scale Correlations

The correlations between the subscales and the Overall IUS score were $r=.970$ for Usable ($p<.001$) and $r=.667$ for Learnable ($p<.001$). The strong scale-to-total correlation for Usable was expected, given that 8 out of 10 items were included in this subscale. The correlation between Usable and Learnable was moderate at $r=.467$ ($p<.001$).

Reliability/Internal Consistency

The overall IUS items had good internal consistency ($\alpha=.83$). Coefficient alphas for Usable and Learnable were $\alpha=.84$ and $\alpha=.67$, respectively. Only two items loaded onto Learnable, contributing to its low alpha. The Learnable subscale was slightly lower than the typical minimum standard of .70 (Landauer, 1997; Nunnally, 1978).

Distribution of IUS Subscale Scores

Table 5 shows descriptive statistics on distributions of scores on the IUS, alongside a comparison with prior SUS data. Of note, the current sample had significantly higher overall scores ($t=3.22$; $p=.001$) and Usable scores ($t=4.11$; $p<.001$) on average, whereas the 2009 SUS study had slightly, non-significantly higher Learnable scores ($t=0.65$; $p=.52$).

Discussion

No psychometrically-sound instruments exist to evaluate the usability of CHIs. Our study of the IUS revealed a factor structure that was nearly identical to Sauro and Lewis (2009). However, our results differed from subsequent studies (Kortum & Sorber, 2015; Lewis et al., 2013; Lewis, Brown, et al., 2015; Lewis, Utesch, et al., 2015; Lewis & Sauro, 2017; Sauro & Lewis, 2011). The moderate correlation between the subscales indicates that the measure can be used as a total scale score, as well as decomposed into Usable and Learnable subscales. The overall IUS score for MI was 68.70, which was slightly below the SUS cutoff of 70 for “acceptable” (Bangor et al., 2008). However, until additional research is conducted with the IUS, it is unknown the extent to which IUS cutoffs translated to CHIs.

Future research should apply the IUS to better understand factors that contribute to high and low scores and the relationships between those scores and related constructs. First, applications with a broader range of EBPI types are indicated to confirm the current results, and these investigations are likely to reveal EBPI qualities that result in higher and lower scores (e.g., EBPI complexity). Second, the IUS should be applied with additional service provider and service recipient populations with a range of experience in the interventions evaluated. Prior research with both the SUS and the IUS suggests that greater experience with products or expertise in domains can result in higher scores (Lyon et al., 2020; Mclellan et al., 2012; Sauro, 2011). Third, initial evidence for the structural validity of the IUS creates opportunities to evaluate the relationships between usability and perceptual implementation outcomes (e.g., acceptability, feasibility) using established measures for those constructs (e.g., Weiner et al., 2017). Fourth, there may be opportunities to improve the IUS itself, such as targeted item developed to enhance the robustness of the two-item Learnable subscale or revised wording including exploration of the utility of including specific examples of EBPIs that demonstrate high or low usability. Finally, future research also should assess the extent to which differences in implementation supports (e.g., training, consultation) impact experiences of EBPI usability.

Limitations

This study had several limitations. First, it occurred only in primary care with one user group and may be difficult to generalize to other contexts. Although the WPRN represents a wide variety of types of health care systems, clinic sizes, and geography (and is broadly representative of primary care clinics in the region), sampling within the WPRN was non-systematic and may have yielded a sample that was more engaged with, or had positive attitudes toward, behavioral health services in general or EBPIs in particular. Second, due to a survey programming error, we did not obtain consistent data about the amount of training respondents had received for MI.

Conclusions And Next Steps

Overall, the adapted IUS demonstrated good psychometric quality and a structure consistent with some prior research. Overall, intervention usability has been conceptualized as a key determinant of both perceptual (e.g., appropriateness, feasibility) and behavioral (e.g., fidelity) implementation outcomes, as well as patient outcomes (Lyon & Bruns, 2019). Application of the IUS to a broader range of EBPIs, settings, and professional roles would allow this proposition to be explicitly tested.

List Of Abbreviations

CHI: Complex health intervention

EBPI: Evidence-based psychosocial intervention

HCD: Human-centered design

IUS: Intervention Usability Scale

MI: Motivational Interviewing

SUS: System Usability Scale

WPRN: WWAMI region Practice and Research Network

WWAMI: Washington, Wyoming, Alaska, Montana, and Idaho

Declarations

Ethics Approval and Consent to Participate

This project was approved by the University of Washington Institutional Review Board (IRB).

Consent for Publication

Not applicable.

Availability of data and material

Please contact the lead author for more information.

Competing Interests

All authors declare that they have no competing interests.

Funding

This publication was supported in part by grants P50MH115837 (Areán) and R34MH109605 (Lyon), award by the National Institute of Mental Health. This publication was also supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1 TR002319. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

ARL developed the Intervention Usability Scale. **ARL, KO, BNR, SAM, and PAA** developed the study survey. **KO** programmed the online survey and served as the liaison to the participating WPRN organizations. **JJ** and **MDP** conducted all data analyses. **ARL** drafted the initial manuscript outline. **ARL, JJ, MDP, and KO** drafted sections of the manuscript. **ARL, JJ, MDP, KO, BNR, SAM, MAA, and PAA** all drafted sections of the manuscript and/or participated in reviewing and approving the final version.

Acknowledgements

We thank Ryan Allred for checking manuscript references and manuscript formatting.

References

- Alexopoulos, G. S., Raue, P. J., Gunning, F., Kiosses, D. N., Kanellopoulos, D., Pollari, C., Banerjee, S., & Arean, P. A. (2016). Engage Therapy: Behavioral Activation and Improvement of Late-Life Major Depression. *American Journal of Geriatric Psychiatry, 24*(4), 320–326. <https://doi.org/10.1016/j.jagp.2015.11.006>
- Areán, P. A., Ayalon, L., Jin, C., McCulloch, C. E., Linkins, K., Chen, H., ... & Estes, C. (2008). Integrated specialty mental health care among older minorities improves access but not outcomes: results of the PRISMe study. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences, 23*(10), 1086-1092.
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction, 24*(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Brooke, J. (1986). *SUS - A quick and dirty usability scale*.
- Burchett, H. E. D., Blanchard, L., Kneale, D., & Thomas, J. (2018). Assessing the applicability of public health intervention evaluations from one setting to another: a methodological study of the usability and usefulness of assessment tools and frameworks. *Health research policy and systems, 16*(1), 88.
- Campbell, N. C., Murray, E., Darbyshire, J., Emery, J., Farmer, A., Griffiths, F., Guthrie, B., Lester, H., Wilson, P., & Kinmonth, A. L. (2007). Designing and evaluating complex interventions to improve health care. *British Medical Journal, 334*(7591), 455–459. <https://doi.org/10.1136/bmj.39108.379965.be>
- Centers for Medicare & Medicaid Services. (2018). *Behavioral Health Integration Services Booklet*. <https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/BehavioralHealthIntegration.pdf>
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2013). Developing and evaluating complex interventions: The new Medical Research Council guidance. *International Journal of Nursing Studies, 50*(5), 585–587. <https://doi.org/10.1016/j.ijnurstu.2012.09.010>
- Dumas, J. S., Dumas, J. S., & Redish, J. (1999). *A practical guide to usability testing*. Intellect books.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations. *The Milbank Quarterly, 82*(4), 581–629. <https://doi.org/10.1111/j.0887-378X.2004.00325.x>
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>
- Hettema, J., Steele, J., & Miller, W. R. (2005). Motivational interviewing. *Annu. Rev. Clin. Psychol., 1*, 91-111. <https://doi.org/10.1146/annurev.clinpsy.1.102803.143833>
- Harte, R., Glynn, L., Rodríguez-Molinero, A., Baker, P. M., Scharf, T., Quinlan, L.R. & ÓLaighin, G. (2017). A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology. *JMIR human factors, 4*(1), e8.
- Horsky, J., Schiff, G. D., Johnston, D., Mercincavage, L., Bell, D. & Middleton, B. (2012). Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. *Journal of biomedical informatics, 45*(6), 1202-1216.

- Institute of Medicine. (2015). *Psychosocial Interventions for Mental and Substance Use Disorders: A Framework for Establishing Evidence-Based Standards*. National Academies Press.
- International Organization for Standardization. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability*.
- International Organization for Standardization. (2019). 9241–210: Ergonomics of human system interaction-Part 210: Human-centered design for interactive systems. Geneva: ISO. <https://www.iso.org/standard/77520.html>
- Kortum, P. T., & Bangor, A. (2013). Usability ratings for everyday products measured with the System Usability Scale. *International Journal of Human-Computer Interaction*, *29*(2), 67-76.
<https://doi.org/10.1080/10447318.2012.681221>
- Kortum, P., & Peres, S. C. (2014). The Relationship Between System Effectiveness and Subjective Usability Scores Using the System Usability Scale. *International Journal of Human-Computer Interaction*, *30*(7), 575–584.
<https://doi.org/10.1080/10447318.2014.904177>
- Kortum, P., & Sorber, M. (2015). Measuring the Usability of Mobile Applications for Phones and Tablets. *International Journal of Human-Computer Interaction*, *31*(8), 518–529. <https://doi.org/10.1080/10447318.2015.1064658>
- Landauer, T. (1997). Behavioral Research Methods in Human-Computer Interaction. In *Handbook of Human-Computer Interaction* (pp. 203–227). North-Holland. <https://doi.org/10.1016/B978-044481862-1.50075-3>
- Lewis, C. C., Fischer, S., Weiner, B. J., Stanick, C., Kim, M., & Martinez, R. G. (2015a). Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. *Implementation Science*, *10*(1), 155.
- Lewis, C. C., Stanick, C. F., Martinez, R. G., Weiner, B. J., Kim, M., Barwick, M., & Comtois, K. A. (2015b). The society for implementation research collaboration instrument review project: A methodology to promote rigorous evaluation. *Implementation Science*, *10*(1). <https://doi.org/10.1186/s13012-014-0193-x>
- Lewis, J. R., Brown, J., & Mayes, D. K. (2015). Psychometric Evaluation of the EMO and the SUS in the Context of a Large-Sample Unmoderated Usability Study. *International Journal of Human-Computer Interaction*, *31*(8), 545–553.
<https://doi.org/10.1080/10447318.2015.1064665>
- Lewis, J. R., & Sauro, J. (2017). Revisiting the Factor Structure of the System Usability Scale. In *Journal of Usability Studies* (Vol. 12). <https://uxpajournal.org/revisit-factor-structure-system-usability-scale/>
- Lewis, J. R., Utesch, B. S., & Maher, D. E. (2015). Measuring Perceived Usability: The SUS, UMUX-LITE, and AltUsability. *International Journal of Human-Computer Interaction*, *31*(8), 496–505.
<https://doi.org/10.1080/10447318.2015.1064654>
- Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE - When there's no time for the SUS. *Conference on Human Factors in Computing Systems - Proceedings*, 2099–2102. <https://doi.org/10.1145/2470654.2481287>
- Lyon, A. R., Brewer, S. K., & Areán, P. A. (in press). Leveraging human-centered design to implement modern psychological science: Return on an early investment. *American Psychologist*. <http://dx.doi.org/10.1037/amp0000652>
- Lyon, A. R. (2016). *Intervention Usability Scale (IUS)*. Unpublished instrument.

- Lyon, A. R., Koerner, K., & Chung, J. (2020). Usability Evaluation for Evidence-Based Psychosocial Interventions (USE-EBPI): A methodology for assessing complex intervention implementability. *Implementation Research and Practice, 1*. <https://doi.org/10.1177%2F2633489520932924>
- Lyon, A. R., & Bruns, E. J. (2019). User-Centered Redesign of Evidence-Based Psychosocial Interventions to Enhance Implementation - Hospitable Soil or Better Seeds? *JAMA Psychiatry, 76*(1), 3-4. <https://doi.org/10.1001/jamapsychiatry.2018.3060>
- Lyon, Aaron R, Munson, S. A., Renn, B. N., Atkins, D. C., Pullmann, M. D., Friedman, E., & Areán, P. A. (2019). Use of Human-Centered Design to Improve Implementation of Evidence-Based Psychotherapies in Low-Resource Communities: Protocol for Studies Applying a Framework to Assess Usability. *JMIR Research Protocols, 8*(10), e14990. <https://doi.org/10.2196/14990>
- Mclellan, S., Muddimer, A., & Peres, S. C. (2012). The Effect of Experience on System Usability Scale Ratings. *Journal of Usability Studies, 7*(2), 56–67. <http://uxpajournal.org/the-effect-of-experience-on-system-usability-scale-ratings/>
- Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change*. Guilford press.
- Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.
- Norman, D. A., & Draper, S. W. (1986). *User Centered System Design; New Perspectives on Human-Computer Interaction*. L. Erlbaum Associates Inc.
- Nunnally, J. C. (1978). *Psychometric theory*. McGraw-Hill.
- Perez Jolles, M., Lengnick-Hall, R., & Mittman, B. S. (2019). Core Functions and Forms of Complex Health Interventions: a Patient-Centered Medical Home Illustration. *Journal of General Internal Medicine, 34*(6), 1032–1038. <https://doi.org/10.1007/s11606-018-4818-7>
- Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: how to plan, design, and conduct effective tests*. Wiley Pub.
- Sauro, J. (2011). *Does prior experience affect perceptions of usability?* <https://measuringu.com/prior-exposure/>
- Sauro, J., & Lewis, J. R. (2009). Correlations among Prototypical Usability Metrics: Evidence for the Construct of Usability. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems: April 4-9, 2009, Boston, MA, USA* (pp. 1609–1618). ACM Press.
- Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? *Conference on Human Factors in Computing Systems - Proceedings, 2215–2223*. <https://doi.org/10.1145/1978942.1979266>
- VanBuskirk, K. A., & Wetherell, J. L. (2014). Motivational interviewing with primary care populations: a systematic review and meta-analysis. *Journal of behavioral medicine, 37*(4), 768-780. <https://dx.doi.org/10.1007%2Fs10865-013-9527-4>
- Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., ... & Halko, H. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science, 12*(1), 1-12. <https://doi.org/10.1186/s13012-017-0635-3>

Tables

Table 1. Study sample demographics.

	N	%
Age		
18-29	37	27.2%
30-39	57	41.9%
40-49	20	14.7%
50-59	12	8.8%
60-69	9	6.6%
Missing	1	0.7%
Gender (self-reported)		
Female/woman	77	56.6%
Male/man	53	39.0%
Genderqueer/non-binary	2	1.5%
Missing	4	2.9%
Race/Ethnicity*		
American Indian/Alaska Native	4	2.9%
Asian	12	8.8%
Black/African American	1	0.7%
Hispanic/Latinx	1	0.7%
Native Hawaiian/Pacific Islander	1	0.7%
White	117	86.0%
Other	3	2.2%
Degree(s) held*		
MD	117	86.0%
DO	10	7.4%
NP	5	3.7%
PA	3	2.2%
MPH	11	5.4%
MS/MA	5	3.7%
BS/BA	41	30.1%
Other	2	1.5%
Years in practice		
0-5	15	11.0%
6-10	11	8.1%
11-20	17	12.5%
21-30	9	6.6%
31 or more	78	5.9%
Missing	60	44.1%

N = 136. *Racial/ethnic categories and degrees held are not mutually exclusive.

Table 2. IUS individual item descriptives for Motivational Interviewing (MI).

Item	Min	Max	Median	Mean	SD
1. I like to use MI frequently.	0	4	3	3.04	0.73
2. I find MI unnecessarily complex.	0	4	1	1.13	0.78
3. I think MI is easy to use.	0	4	3	2.54	0.90
4. I need the support of an expert consultant to be able to use MI.	0	3	1	0.88	0.80
5. I find the various components of MI are well integrated.	1	4	3	2.72	0.73
6. I think there is too much inconsistency in MI.	0	4	1	1.19	0.81
7. I would imagine that most people would learn to use MI very quickly.	0	4	3	2.38	0.97
8. I find MI very cumbersome to use.	0	4	1	1.11	0.88
9. I feel very confident using MI.	1	4	3	2.55	0.82
10. I needed to learn a lot of things before I could get going with MI.	0	4	1	1.43	0.92

N = 136 for all items. For all items, 0 = strongly disagree, 1 = disagree; 2 = neither agree nor disagree; 3 = agree; 4 = strongly agree. Items in bold are reverse scored.

Table 3. Item Correlations

	1. I like to use MI frequently.	2. I find MI unnecessarily complex.	3. I think MI is easy to use.	4. I need the support of an expert consultant to be able to use MI.	5. I find the various components of MI are well integrated.	6. I think there is too much inconsistency in MI.	7. I would imagine that most people would learn to use MI very quickly.	8. I find MI very cumbersome to use.	9. I feel very confident using MI.	10. I needed to learn a lot of things before I could get going with MI.
1. I like to use MI frequently.	-	-	-	-	-	-	-	-	-	-
2. I find MI unnecessarily complex.	-0.37***	-	-	-	-	-	-	-	-	-
3. I think MI is easy to use.	0.30***	-0.52***	-	-	-	-	-	-	-	-
4. I need the support of an expert consultant to be able to use MI.	-0.28**	0.40***	-0.38***	-	-	-	-	-	-	-
5. I find the various components of MI are well integrated.	0.33***	-0.42***	0.51***	-0.34***	-	-	-	-	-	-
6. I think there is too much inconsistency in MI.	-0.05	0.40***	-0.26**	0.31***	-0.37***	-	-	-	-	-
7. I would imagine that most people would learn to use MI very quickly.	0.10	-0.41***	0.58***	-0.28**	0.33***	-0.19*	-	-	-	-
8. I find MI very cumbersome to use.	-0.35***	0.64***	-0.56***	0.35***	-0.38***	0.48***	-0.48***	-	-	-
9. I feel very confident using MI.	0.45***	-0.41***	0.56***	-0.46***	0.43***	-0.16	0.47***	-0.44***	-	-
10. I needed to learn a lot of things before I could get going with MI.	-0.10	0.20*	-0.18*	0.50***	-0.21*	0.15	-0.27**	0.25**	-0.29**	-

N = 136. *Significant at p < .05; **Significant at p < .01; ***Significant at p < .001.

Table 4. Rotated Component Matrices for three exploratory structure analyses

Item	Principal axis		Principal components		Unweighted least squares	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
	1. I like to use MI frequently.	0.383	-0.210	0.412	-0.293	0.387
2. I find MI unnecessarily complex.	-0.702	0.234	-0.771	0.182	-0.698	0.238
3. I think MI is easy to use.	0.769	-0.160	0.785	-0.178	0.766	-0.164
4. I need the support of an expert consultant to be able to use MI.	-0.273	0.843	-0.305	0.773	-0.237	0.970
5. I find the various components of MI are well integrated.	0.567	-0.225	0.649	-0.205	0.572	-0.213
6. I think there is too much inconsistency in MI.	-0.402	0.192	-0.561	0.041	-0.398	0.198
7. I would imagine that most people would learn to use MI very quickly.	0.585	-0.169	0.622	-0.218	0.592	-0.156
8. I find MI very cumbersome to use.	-0.748	0.191	-0.797	0.160	-0.748	0.189
9. I feel very confident using MI.	0.578	-0.366	0.568	-0.486	0.589	-0.341
10. I needed to learn a lot of things before I could get going with MI.	-0.159	0.535	-0.028	0.867	-0.185	0.469

N=136

Table 5. Comparison with prior System Usability Scale data.

Statistic	Lewis & Sauro SUS (comparison sample)			Current Dataset		
	Overall	Usable	Learnable	Overall	Usable	Learnable
<i>N</i>	324	324	324	136	136	136
<i>Mean</i>	62.10	59.40	72.70	68.70**	68.15***	71.14
<i>Variance</i>	494.38	531.54	674.47	183.29	203.40	346.10
<i>Standard Deviation</i>	22.24	23.06	25.97	13.54	14.26	18.60
<i>Standard Error</i>	1.24	1.28	1.44	1.17	1.23	1.60
<i>Skewness</i>	-.42	-.38	-.80	-.08	-.16	-.39
<i>Kurtosis</i>	N/A	N/A	N/A	-.20	-.09	-.30

Significantly different from comparison sample at $p = .01$;* "" at $p = .001$.