UC Riverside

UC Riverside Previously Published Works

Title

Development of the Usage Rating Profile-Web Resource (URP-WR): Using Assessment to Inform Web Resource Selection

Permalink

https://escholarship.org/uc/item/7xc3122f

Journal

Computers in the Schools, 37(4)

ISSN

0738-0569

Authors

Mandracchia, Nina R
Sims, Wesley A

Publication Date

2020-10-01

DOI

10.1080/07380569.2020.1835388

Peer reviewed

Development of the Usage Rating Profile-Web Resource (URP-WR): Using Assessment to Inform Web Resource Selection

Nina R. Mandracchia & Wesley A. Sims

Published online: 27 Oct 2020.

Submit your article to this journal ↗

Article views: 70

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

Check for updates

Development of the Usage Rating Profile-Web Resource (URP-WR): Using Assessment to Inform Web Resource Selection

Nina R. Mandracchia and Wesley A. Sims

University of California, Riverside, California, USA

ABSTRACT

As technology use continues to rapidly increase, so too does consumer use of web-based resources. While important, accessibility is often overemphasized by users when consuming and evaluating web resources. This prioritization may have particularly negative consequences for the selection of supports or interventions in educational settings. This study outlines the development of Usage Rating Profile-Web Resource (UPR-WR), an assessment designed to support the identification, evaluation, and use of web resources. Additionally, preliminary validity evidence supporting the URP-WR interpretation and use argument (IUA) is presented. Content validation and exploratory factor analyses results yielded an URP-WR comprised of 31 items across four identified factors: plausibility, accessibility, appearance, and systems support. Extracted factors demonstrated acceptable reliability estimates ($\alpha = 0.82$–0.93) and levels of social validity. This resulting URP-WR stands to provide assessment data that will aid educators in selecting web resources of quality for use in a practical setting.

Use of evidence-based practices and interventions (EBP/I) is critical to promoting success for all students (Hunsley, 2007; Raines & Raines, 2008). Advances in technology and Internet accessibility have increased exponentially in the last two decades (Fischer-Baum, 2017), resulting in educators' use of web resource informed practices (Cummings, 2011). Search engines make it possible to easily access and disseminate information. To bridge the apparent gap between research and practice, some scholars and educators have created web resources that collect, summarize, and organize EBP/ I to support their dissemination. Universities, state department of education agencies, federal agencies, educators, scholars, professional organizations, and businesses are increasingly organizing information, resources, and products onto web-based platforms to support training and implementation

of evidence-based practice (e.g., ies.gov/ncee/wwc, interventioncentral.org, nasponline.org, intensiveintervention.org, ebi.missouri.edu). Unfortunately, the volume of available information coupled with the possibility of low-quality information may overwhelm educators and negatively impact their judgments around usability of available web-based information (Buchanan & Knock, 2001). In short, too much information too quickly can be overwhelming, confusing, or distracting, resulting in less objective utilization decisions. Internet usage analysis for web resources suggests that 53% of users select the top result (Miller, 2012). While intuitively the first option might appear to be the best option, a variety of factors unrelated to the quality or credibility of the resource may influence the order of search results. Search engine optimization (SEO) allows content developers to align their content with commonly searched keyword phrases (Pinkerton, 2000). While an important consideration when consuming and using information, ease of access should not be the only consideration as accessibility may not reflect quality or usability. Rosenfield (2000) defined usability as a multifaceted process by which knowledge is disseminated in ways that truly inform a consumer. Objective, data-based decision-making driven by a defensible, feasible evaluation of web-resource usability offers a potential solution. This study presents the initial development and preliminary psychometric evidence for the User Rating Profile-Web Resource (URP-WR), an assessment of web resource usability designed to facilitate objective consumer evaluation and comparison of web resources.

Internet information quality

The advent of social media, internet marketing, and easily accessible web-page building and hosting services have created the opportunity for easy dissemination of low-quality, non-evidence-based information, whether or not well-intentioned. Though unexplored in education, the medical field has documented several instances of dissemination of poor quality or blatantly incorrect information via web resources (Benotsch, Kalichman, & Weinhardt, 2004; Morr, Shanti, Carrer, Kubeck, & Gerling, 2010; Ream, Blows, Scanlon, & Richardson, 2009). Given educator reports of consumption preferences (see Johnson et al., 2019), similar findings are likely to extend to the field of education. The abundance of information, coupled with the potential for it to be of low quality or incorrect, puts a well-intentioned educator in a challenging position. Without guidance needed to objectively evaluate internet resources, information overload may result in consumer use of resources based solely on ease of access.

Beyond accessibility, Schrock (2019) outlined additional factors users should consider when evaluating web resources, including credibility,

authority, accuracy, objectivity, currency, coverage, and appearance. Additionally, efficiency and verifiability are noted as considerations for web resource usability (Lydia M. Olson Library, 2018). Credibility, authority, accuracy, verifiability, and objectivity all relate to the information on which the resource is based. These considerations speak to the importance of the reliability and integrity of the information in the resource. In short, this is about whether the evidence underlying the resource is sound and derived from reliable sources. Appearance and efficiency address how the resources present information and how easily a user can access and use it. Coverage refers to the depth and breadth of the resource's information. Collectively, the presence of these attributes equates to a high-quality, usable web resource.

Informed decision-making

Digital citizenship refers to a construct that includes protection of information, detection of information quality, and safe and responsible internet use (e.g., cyberbullying, photo/video exchange, social media use; Ribble, 2012). Available literature suggests that many professions, including education, largely neglect training in digital citizenship (Isman & Canan Gungoren, 2014; Ribble & Bailey, 2004). Illustrating this point, Johnson et al. (2019) found special educators reported a variety of issues with finding quality information, including not knowing where to look, changing terminology, and lack of access to scholarly journals. As a result, these educators reported frequent use of practices disseminated via sources that lack oversight or quality control measures (e.g., Pinterest, Facebook, Instagram; Johnson et al., 2019). Ideally, high-quality practices or procedures would include an empirical evidence base while also being feasible and easily accessible. Unfortunately, the absence of one often renders a resource useless in spite of the presence of others. An evidence-based practice or intervention (EBP/I) that is inaccessible is of little value to users, while a feasible intervention that lacks an evidence base (i.e., evidence of efficacy) will likely fail to address the identified need (Noell & Gansle, 2014).

One potential solution for under-developed digital citizenship lies in the use of objective assessment data to guide consumption decisions. Underlying the collection and use of data for decision making is the belief that resulting decisions are less subjective, less biased, and more equitable (Hintze, Volpe, & Shapiro, 2002; McIntosh, Barnes, Eliason, & Morris, 2014; Scott, Hirn, & Barber, 2012). Application of data-based decision-making principles applied to web resources could result in making more informed subjective evaluations of these resources. Ultimately, these evaluations would lead to adoption of higher quality resources, those that are not

only accessible but also evidence-based. With this in mind, an assessment designed to facilitate objective, data-based consumer decisions about web resources appears advantageous. Unfortunately, available assessments to support data-based decision making through evaluating the usability of web-based educational resources are limited.

Currently available evaluations of web resources

A review of publicly available web-resource evaluation tools yielded few options. The limited available tools for evaluating web resources are frequently found on university websites (see Lydia M. Olson Library, 2018; Schrock, 2019). The Lydia M. Olson Library (2018) at Northern Michigan University provides students with questions to use when evaluating internet sources across six criteria. Similarly, Schrock (2019) provided a number of checklists for evaluating educationally inclined web-based resources.

While well-intentioned, several challenges are evident for these evaluations. Currently available evaluations include questions formatted in a dichotomous (yes/no) manner consistent with a checklist. Although such formatting can be helpful, it does not provide information that facilitates low-inference evaluations or comparisons (Christ, Riley-Tillman, & Chafouleas, 2009; Greenwald & O'Connell, 1970). Other available evaluation tools seem lengthy, making their use complicated. While these assessments represent good faith attempts to evaluate web resources, the noted shortcomings of these tools suggest the need for development of a psychometrically sound, objective, and usable assessment of web resources.

URP-WR

To address the limited availability of measures by which the quality of web-based resources can be evaluated, this study begins the development, refinement, and validation process for the URP-WR. The URP-WR was designed to objectively evaluate web resources across four hypothesized domains: accessibility, appearance, credibility, and feasibility. The URP-WR extends the usage rating profile (URP) formatting and assessment methodology, an approach used to evaluate and inform user perceptions and adoption of educational intervention and assessment practices, to web resources (see Chafouleas, Miller, Briesch, Neugebauer, & Riley-Tillman, 2012; Chafouleas et al., 2009).

Development of the URP-WR is best conceptualized as a development *and* validation process, as development often occurs while anticipating a series of steps to accumulate validity evidence (Kane, 2013). The interpretation/use argument (IUA) approach to assessment validation begins by

clearly stating the proposed interpretations and uses of the assessment (Kane, 2013, p. 2). Next, the inferences made in the IUA are evaluated by empirically testing stated claims (Cook, Brydges, Ginsburg, & Hatala, 2015). Inferences underlying assessment validation are defined by Kane as (a) scoring, or the assigning an accurate, reproducible quantifiable score to an observation; (b) generalizing, or obtaining a representative sample so that the assessment can be applied across multiple possible scenarios; (c) extrapolation, or interpreting the scores from the representative sample as applying to real-life scenarios validly; and (d) implication, or making a decision based on the interpretation (Cook, Brydges, Ginsburg, & Hatala, 2015; Kane, 2013).

This study

The goals of this study were to (a) develop an instrument that will inform and guide consumer use of web resources and (b) increase likelihood of EBP/I implementation in schools, effectively decreasing the persistent research to practice gap in education (Cummings, 2011; Flaspohler, Anderson-Butcher, Paternite, Weist, & Wandersman, 2006; Greenwood & Abbott, 2001). The URP-WR will also serve to inform web resource developers as they endeavor to create high-quality web resources. This study presents (a) initial development processes and (b) refinement processes for the URP-WR. This work addressed assumptions underlying the scoring inferences for the URP-WR through initial content validation and reliability evaluation. Specific research questions included the following:

1. Will a hypothesized factor structure hold (i.e., research suggested usability comprised of four components)?
2. What item combinations are the most appropriate to maximize URP-WR utility (i.e., valid results using the fewest number of items possible)?
3. Does the URP-WR demonstrate reliability through internal consistency of items within factors?
4. Do users perceive this measure as usable (i.e., social validity)?

Materials and methods

Participants

This study occurred across two broad stages: (a) initial item development and content validation and (b) pilot administration. Participants were recruited through email, social media, in-class announcements, and instructors.

Table 1. Demographic characteristics of participants.

Category	Subcategory	*n* EFA	*N* URP-A
Gender	Female	76	62
	Male	18	13
Ethnicity	Hispanic	38	30
	White (not Hispanic)	25	22
	Asian or Pacific Islander	15	12
	Biracial/Multiracial	9	7
	Middle Eastern	4	4
	Other	2	1
Age	18–22	42	30
	23–29	28	24
	30+	21	21
Area of study (students)	Education	50*	38*
	Psychology	12	5
	Other	15	10
Job function (professionals)	Teacher	20*	19*
	Aide/Assistant	5	3
	Professor	3	1
	Other	28	20

Note. Some participants identified as both students (marking an area of study) and teachers.

Participants in the development and initial content validation phase
Participants were recruited from the University of California, Riverside (UCR) faculty and students in the school psychology program. Eight participants completed the consensus-building task. These included six students and two faculty members in the school psychology program.

Participants in the factor analysis and item (data) reduction phase
Participants included 94 faculty, in-service educators, and undergraduate and graduate students in the fields of education and psychology. Participants were recruited through emails and announcements in classes, research lab meetings, and social media. The majority of participants were female ($n = 76$) and Hispanic ($n = 38$). The majority of participants were students ($n = 62$) studying education ($n = 50$). There were also a fair number of teachers ($n = 20$). The average age of participants was 29, but the majority of participants fell in the 18–22 age range ($n = 42$). See Table 1 for demographic characteristics of participants. Note that not all participants responded to every question; thus, some of the numbers are lower than expected. Additionally, note that students are counted under "area of study" and thus not counted under "job function," as some participants have roles as both students and teachers. Thus, these numbers are higher than the overall number of participants.

Participants in the social validity measure phase
A total of 75 participants elected to participate in the social validity part of the study, as an option was provided with additional incentive (i.e., more

raffle entries for a chance to win a gift card). The demographic characteristics largely matched those in the data reduction portion.

Measure

Measure 1: URP-WR

The initial URP-WR included 70 items organized across four hypothesized domains encompassing usability. Items were formatted as statements capturing the aspects of usability across appearance, accessibility, credibility, or feasibility. For example, "The resource cites its original sources." Raters responded to statements using a 6-point Likert-like scale (1 = *strongly disagree* to 6 = *strongly agree*). This formatting is consistent with that of the other available URP tools.

Measure 2: Usage Rating Profile-Assessment (URP-A)

The URP-A is a 28-question evaluation of social validity of an assessment. Social validity refers to the subjectively measured acceptability of procedures, in this case filling out the URP-WR, and users' satisfaction with the instrument (McNeill, 2019; Wolf, 1978). The URP-A has items formatted on a 6-point Likert-like scale (1 = *strongly disagree* to 6 = *strongly agree*) . The URP-A has a six-factor model with acceptable model fit. The alpha coefficients ranged from .71 to .90, having acceptable reliability with the exception of one factor (system support α = .63; Miller, Neugebauer, Chafouleas, Briesch, & Riley-Tillman, 2013).

Procedures

The development and initial validation followed a four-step process. First, IUA development of the URP-WR was initiated, with scoring being the inference tested in this study (Kane, 2013). Preliminary item development and then preliminary content validation in the form of the consensus-building task were completed. Finally, the initially developed items were administered with a provided web-based resource in order to collect data for an exploratory factor analysis.

URP-WR IUA

The IUA for the URP-WR is to effectively evaluate web-based resources promoting EBP/I for usability of the resource itself as well as implementation feasibility of the proposed recommendations based on a user's perceptions. In other words, if the user perceives the resource to have high levels of accessibility, pleasing appearance, credibility, and feasibility, the URP-WR score would be high. This is important because it gives a quantifiable

score associated with a user's perception that can be used to make an informed decision about selection of resources.

Development

The development of the URP-WR began with a literature review to determine the factors underlying a quality web-based resource. The scarcity of peer-reviewed research on characteristics underlying usable web-based resources necessitated inclusion of non-peer-reviewed, publicly available information in this review. Characteristics were drawn largely from a variety of resources including those provided by Kathy Schrock, university library Web sites, and existing URP assessment tools. Identified trait or characteristic considerations for usable web-based resources were grouped by commonality into broad themes. Noted commonalities resulted in identification of four usability domains:

1. **appearance**: Characteristics consistent with appearance included visual appeal, organization, use of pictures, use of headings, use of advertisements, size of font, and more. These characteristics were combined to encompass *appearance*, which includes the esthetic appeal as well as logical organization of the resource.
2. **accessibility:** Characteristics consistent with accessibility included the ease of finding the resource, ease of using the resource, length of time needed for the resource to load, the presence of different modalities (e.g., option to read or listen to the information presented), the presence of cost associated with accessing the resource, and more. These characteristics were combined to encompass *accessibility*, which includes the ease associated with accessing and utilizing the resource.
3. **credibility:** Characteristics consistent with authorship and credibility included citations, date of citations, name recognition of the author, the presence of bias in the citations, availability of the author for contact, and more. These characteristics were combined to encompass *credibility*, which takes into account citations and links as opposed to just the authority of the author.
4. **feasibility:** Characteristics consistent with feasibility were need for administrative support, need for consultative support, the amount of time it would take to implement the recommendations provided in the resource, and more. These characteristics were combined to encompass *feasibility*, which includes the practicality associated with implementing the recommendations provided in the resource.

Preliminary item development

Next, items were generated relative to this literature review. Development and formatting were largely modeled after those employed when developing

other URP assessments (e.g., Chafouleas et al., 2009). Initially, 70 items emerged from previous resources including the URPs, Schrock, and others for initial content validation (Chafouleas et al., 2009; Schrock, 2019).

URP-WR content validation

Preliminary URP-WR construction

Initial content validation of a hypothesized four-factor structure was conducted in the form of a consensus-building task (Hennessy et al., 2016). The author constructed definitions for these four theoretical components using the limited web-based resource literature (e.g., Schrock, 2019). To test these theoretical groupings, nine UCR students and faculty individually rated the 70 items on a scale from 1 to 4 (1 = *best fit*) indicating fit within the four hypothesized factors. For example, the item "this resource was easy to access" would hypothetically show best fit with "accessibility," so the participant would mark 1 for accessibility on this item, potentially 2 for appearance, and so on. Given their prior knowledge and experience in assessment, research, and technology utilization, these participants were considered appropriate judges to sort items in a logical and informed manner. Analysis of rater consensus was used to inform further item reduction.

URP-WR factor analysis and item (data) reduction

Once developed, a pilot study was conducted using the drafted URP-WR through Qualtrics. Participants first completed the informed consent form, then viewed the video instructions, and read the written instructions. Completion of the URP-WR included three steps: (a) conduct a Google search, (b) open a link to a provided web resource, and (c) use the Google search and provide a web resource to rate the statements in the URP-WR.

The draft URP-WR included 55 items organized across four hypothesized factors. Upon completion of the URP-WR, participants were given the option to continue on to complete the URP-A and earn two extra entries into the gift card drawing or to stop after completion of the URP-WR and earn one entry. This method was chosen in order to reduce attrition.

Social validity

Social validity refers to the subjectively measured acceptability of procedures, in this case filling out the URP-WR, and users' satisfaction with the instrument (McNeill, 2019; Wolf, 1978). Raters who elected to participate used the URP-A to evaluate the acceptability of the URP-WR as an assessment tool itself (Chafouleas et al., 2012). It was made clear that participants

were expected to evaluate their satisfaction with the process of completing URP-WR and not the web resource provided. The inclusion of this measure provided a measure of social validity of the URP-WR.

Results

Research question 1

The first research question sought to determine whether an initially hypothesized four-factor structure would emerge. Initially, a consensus-building task was conducted to reduce items. Fifteen items (items 4, 12, 17, 19, 20, 23, 27, 28, 29, 41, 51, 55, 57, 61, and 63) did not meet a cutoff of 75% agreement on category of best fit. Thus, 55 items remained and were included in the factor analysis.

Data were examined to ensure appropriateness for factor analysis. The assumption of ratio level of measurement was met due to the use of a Likert-like scale that functioned as a continuous variable in analyses. The assumption of random sampling was violated because of the use of a convenience sample. This is common for pilot studies, but still results should be interpreted with caution. The assumption of normality was violated as the data were largely not normally distributed. This was accounted for through use of the OLS method of extraction, which is also suggested because of the small sample size (Briggs & MacCallum, 2003). The assumption of linear relations between observed variables was met as the bivariate scatterplots demonstrated linear relationships.

All 55 items met criteria of showing significant correlation ($r \geq .30$) with at least two items as well as not showing multicollinearity with at least three other items. Therefore, no items were deleted due to either multicollinearity (redundancy) or a lack of correlation (unnecessary items). Finally, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy demonstrated a meritorious value, KMO = 0.81 (Kaiser, 1974). Thus, the data were deemed ready and suitable for factor analysis.

The number of factors to extract was chosen based on a parallel analysis, scree plot, as well as interpretability as determined by the researcher. The break in the scree plot demonstrated that between four and six factors should be extracted. The parallel analysis supported this and suggested extraction of four factors. Eigenvalues indicated that 55.22% of the variance in the data was explained using four factors. When additional factors were extracted, no items met the decision rules within the extra factors. Therefore, although additional variance is explained through additional factors, four factors were extracted to eliminate redundancy for interpretability and because four factors were suggested by the parallel analysis and scree plot.

The results of the EFA demonstrated a root-mean-square of residuals (RMSR) value that was close to 0 as expected (RMSR = 0.05). Additionally, the EFA fit value was close to 1 as expected (fit based upon off-diagonal values = 0.97). Therefore, the fit statistics demonstrated that this model showed good fit. A three-factor EFA (fit based upon off-diagonal values = 0.96, RMSR = 0.06) and five-factor EFA (fit based upon off-diagonal values = 0.98, RMSR = 0.05) were also conducted but were determined to not fit the data as well as the four-factor model.

Research question 2

A second research question sought to find what item combinations were appropriate to maximize URP-WR utility. Final items were retained based on pattern loading (see Table 2). Based on decision rules, 34 items were retained. Those items are bolded for ease of interpretation. Estimates of communality from retained items ranged from 0.33 to 0.78. As anticipated, factor correlations. See Table 3 for Factor Correlations.

Additionally, a second EFA was conducted using the items retained to ensure the factor structure still held with the remaining items. This EFA also demonstrated acceptable fit statistic levels (fit based upon off-diagonal values = 0.97, RMSR = 0.06). The four-factor structure emerged once again according to parallel analysis and a scree plot. After this EFA, three items (items 34, 46, and 66) did not meet the decision rule of loading of at least 0.45 on their primary factor and loading no larger than 0.30 on a secondary factor and were thus eliminated.

Finally, the last EFA was conducted on the 31 remaining items. The four-factor structure emerged once more according to parallel analysis and a scree plot. The four-factor structure explained 55.77% of the variance in the data. The final EFA had acceptable fit statistic levels (fit based upon off-diagonal values = 0.98, RMSR = 0.05). All items fell within the decision rules and thus were retained in the final version of the URP-WR.

Research question 3

Estimates of reliability within the four factors were calculated using Cronbach's alpha. All four factors demonstrated acceptable levels of internal reliability. Factor I contained 12 items and demonstrated a high level of internal reliability ($\alpha = 0.93$, 95% CI = 0.90–0.95). Factor II contained 10 items and demonstrated a high level of internal reliability ($\alpha = 0.90$, 95% CI 0.87–0.93). Factor III contained four items and demonstrated an acceptable level of internal reliability ($\alpha = 0.84$, 95% CI

Table 2. Pattern and structure factor loadings of initial exploratory factor analysis.

Item no.	Item	Factor I		Factor II		Factor III		Factor IV		h^2
		P	S	P	S	P	S	P	S	
Factor I: Plausibility (12 of 26 items retained)										
38	The information is from sources known to be reliable.	**0.88**	0.83		0.27		0.13		0.28	0.70
22	The resource provides citations from reliable sources.	**0.79**	0.84		0.35		0.21		0.39	0.71
37	The resource provides citations.	**0.71**	0.71				0.18		0.37	0.53
44	This resource appropriately represents the context of its cited sources.	**0.69**	0.68						−0.27	0.54
45	The sources used by the resource provided appear credible.	**0.68**	0.79		0.30	0.19	0.38		0.42	0.67
54	I understand the components of the recommendations provided in this resource.	**0.65**	0.74						−0.17	0.58
31	I would know what to say if I were asked how to implement the recommendations provided in this resource.	**0.64**	0.67		0.26		0.34		0.22	0.49
70	I believe information from this resource.	**0.63**	0.77	0.27	0.50		0.28		0.34	0.67
18	The resource contains all recommendations needed for implementation.	**0.59**	0.74			0.19				0.57
10	The resource cites its original sources.	**0.59**	0.53			−0.24			0.22	0.33
35	The resource appropriately cites ideas that were not its own.	0.58	0.66		0.30		0.25		0.32	0.46
36	The cost of implementing recommendations in this resource is reasonable.	0.58	0.71	0.28		0.17	0.25	0.21	0.32	0.59
67	I am convinced that the resource is accurate as a direct result of appropriate citations.	0.56	0.75	0.41	0.62		0.20		0.34	0.72
49	I would feel confident sharing these recommendations with my colleagues.	0.55	0.72	0.31	0.51	0.22	0.36		0.31	0.64

Item	Statement								
53	Information for original resource sources are easily identifiable.	**0.54**	0.60	0.23	0.41	0.17		0.16	0.37
64	I would be willing to implement this resource's recommendations in my setting.	0.51	0.67	0.37	0.55	0.28	0.16	0.22	0.58
5	The information appears to be valid and well-researched.	0.50	0.65	0.33	0.51			0.30	0.51
16	Topics are successfully addressed, with clearly presented arguments and adequate support to substantiate them.	**0.49**	0.61	0.16	0.35	0.17	0.16	0.37	0.42
24	The recommendations could be feasibly implemented in my setting.	0.49	0.64	0.30	0.47	0.36	0.25	0.22	0.53
50	I understand how to implement the recommendations described in this resource.	0.48	0.55		0.16	0.57	0.49		0.52
25	The amount of time required to effectively implement the recommendations provided in this resource is reasonable.	0.46	0.67	0.26	0.43	0.40	0.26	0.40	0.58
9	This resource provides citations in proper APA or another format.	0.46	0.50			−0.25	0.32	0.44	0.38
3	The recommendations provided in this resource could easily be implemented as described.	0.44	0.64	0.29	0.28	0.36	0.20	0.51	0.53
7	I was able to save this resource for future use.	0.40	0.55	0.16	0.22	0.16	0.31	0.48	0.39
47	There were no 404 errors or others that blocked me from accessing this resource.	0.31	0.24	−0.23			0.11		
43	Quotes and other strong assertions are backed by sources that one could check through other means.	0.39	0.72	0.24	0.39			0.25	0.30
Factor II: Appearance (10 of 13 items)									
60	The design of the resource makes me more likely to use it.	0.32	**0.89**	0.88			0.78		
56	This resource looks appealing.	0.33	**0.86**	0.86				0.16	
65	I wish more resources were designed the way this one is.	0.37	**0.83**	0.85			0.73	0.16	0.75

(continued)

Table 2. Continued.

Item no.	Item	Factor I		Factor II		Factor III		Factor IV		h^2
		P	S	P	S	P	S	P	S	
11	This resource is esthetically pleasing.	-0.16		**0.73**	0.68				0.43	
69	The site appears well maintained.	0.29	0.51	**0.66**	0.76			0.18	0.65	
58	The resource was updated recently enough for me to trust it.	0.16	0.40	**0.65**	0.71			0.52		
42	This resource looks professional.	0.23	0.43	**0.60**	0.69			0.51		
21	Pictures or photographs in the resource add to the information.			**0.57**	0.55	-0.22	-0.25		0.38	
68	This resource looks well organized.	0.36	0.63	0.50	0.64	0.16	0.27	0.33	0.34	0.63
59	I was able to download this document as a Word doc or PDF for future use.	0.20	0.32	**0.47**	0.54					
52	There is an image map (large clickable graphic with hyperlinks) on the resource.			**0.47**	0.43	-0.16	-0.22	-0.26	-0.26	0.36
6	This resource was easy to use.	0.34	0.64	0.34	0.50		0.30	0.38	0.57	0.65
39	There are helpful headings and subheadings on the resource.	0.33	0.58	0.30	0.44	0.15	0.29	0.27	0.46	0.49
Factor III: System Support (6 of 7 items)										
30	I would need support from my administrator to implement recommendations made in this resource.					**-0.83**	0.78			0.65
32	Support from administration would be needed to implement recommendations provided in this resource.			-0.22	-0.26	**0.79**	0.79	0.20	0.20	0.70
8	I could only implement recommendations in this resource with assistance from other adults.					**0.68**	0.65			0.43
46	I could implement the recommendations in this resource by myself.	0.28	0.37			**0.60**	0.65		0.15	0.48
34		0.22	0.39			**0.49**	0.57		0.33	0.36

Item	Description									
	I have the skills needed to implement the recommendations provided in this resource.									
13	Implementation of the recommendations made in this resource would require support from my coworkers.		0.16			**0.41**	0.45	0.25	0.60	0.24
48	This resource ignores important elements from its cited sources.		0.35		0.17	−0.35	−0.43	−0.22	−0.36	0.30
Factor IV: Accessibility (6 of 9 items)										
15	It was easy to find this resource from a simple Google search.		0.28					**0.84**	0.82	0.68
1	The resource was easy to find.		0.27			−0.17		**0.76**	0.74	0.58
26	It was easy to find this resource.		0.43					**0.80**	0.84	0.73
66	This resource is easily accessible.		0.46	0.21	0.31		0.17	**0.71**	0.77	0.66
2	It was difficult to find this resource from a simple Google search.	−0.28			0.31			**0.58**	0.49	0.31
40	This resource required too many links to find.		0.36			0.19	0.34	**0.63**	0.70	0.53
33	It was easy to find this resource without guidance.	0.17	0.41		0.20		0.26	0.44	0.54	0.36
14	This resource took a long time to load.	0.30	0.44				0.23	0.39	0.52	0.35
62	I am satisfied with the amount of time this resource took to load.	0.40	0.59		0.24		0.19	0.42	0.58	0.50

Note. Coefficients below 0.15 were suppressed. Items that meet decision rules are in bold.

Table 3. Factor correlations.

Subscale	Factor I	Factor II	Factor III	Factor IV
Plausibility	1.00			
Appearance	0.35	1.00		
System support	0.24	0.01	1.00	
Accessibility	0.39	0.09	0.21	1.00

Table 4. Reliability statistics for factors.

Factor	Average inter-item r	SD of inter-item r	α	95% CI (α)
Plausibility	0.52	0.01	0.93	0.90–0.95
Appearance	0.49	0.02	0.90	0.87–0.93
System support	0.57	0.03	0.84	0.79–0.89
Acceptability	0.51	0.03	0.82	0.77–0.88

Table 5. Social validity scores: URP-A averages by category.

Category	Average item score	Average score
Acceptability	4.27	29.89
Understanding	4.46	13.37
Feasibility	4.55	27.31
System climate	4.24	16.95
System support	3.73	11.19

Note. Categories that met the goal score.

0.79–0.89). Factor IV contained five items and demonstrated an acceptable level of internal reliability ($\alpha = 0.82$, 95% CI 0.77–0.88; see Table 4).

Research question 4

Usability was assessed through administration of the URP-A. Of the 94 participants that completed the URP-WR pilot study, 75 elected to continue and answer the URP-A questions. The best overall score possible on the URP-A would be 138 (23*6). The goal score was 92 (23*4). The overall average URP-A score across participants was 98.61 (see Table 5). This indicated that the users perceived the URP-WR to be socially valid and acceptable to use in their setting. Additionally, average scores were calculated per category using the same criteria. All categories except systems support met the category goals.

Discussion

This study sought to provide initial support for and development of the URP-WR. The consensus-building task demonstrated initial content validity of the URP-WR. This falls under the scoring inference of the IUA. Participants sorted the items into four categories, and items that did not reach 75% agreement on one category as best fit were eliminated. This resulted in the elimination of 15 items, as well as initial content validation of the URP-WR.

The results of the EFA demonstrated that four factors should be extracted. The EFA demonstrated acceptable levels of fit (0.97) and RMSR (0.06), and thus can be interpreted reliably. This factor structure explained just over half of the variance found in the data. It is satisfactory in social science research to see 50% to 60% of the variance explained through EFA; thus, this is satisfactory (Hair, Black, Babin, & Anderson, 2010). Subsequent factor analyses demonstrated similar levels of fit, explanation of variance, as well as the same numbers of factors extracted. However, the four factors extracted were somewhat different from the hypothesized factors of accessibility, appearance, credibility, and feasibility. Rather, two factors, accessibility and appearance, emerged as hypothesized, while items from credibility and feasibility emerged as one factor, and systems support emerged on its own.

The factor accessibility and appearance emerged essentially as expected. Items were retained in those factors to measure these aspects of the web-based resource itself. These factors were shown to be important to the evaluation of web-based resources in the literature review. Thus, it makes sense that they emerged separately and significantly in the factor analysis.

Many of the items initially hypothesized to make up the factors of feasibility and credibility ultimately emerged as one factor. These items focused on the citations and believability of the information as well as feasibility of the recommendations provided in the resource. Thus, this factor was relabeled *plausibility*. A high score on the plausibility scale indicated that the user perceived this resource as containing information from credible sources that can be easily understood and implemented practically. The items related to credibility had higher factor loadings than those relating to feasibility in general. However, both of these aspects emerged as important to web-based resources in the literature review. Thus, both were included in the final factor that was renamed plausibility. This word was chosen as it encompasses the plausibility of implementation as well as the plausibility of the research base supporting the resource. This result was somewhat unexpected, as the research literature did not establish the connection between feasibility of the recommendations and the credibility of them. In fact, this may suggest that the two may be at odds which made this result surprising. It is possible that this connection arose because of the resource used. Participants saw the evaluated WBR as containing credible sources and having feasible recommendations. Thus, the EFA did not pull the two factors apart as expected. This should be empirically evaluated further.

The second factor contained items focused on the overall design and appeal of the resource. Thus, this factor was labeled *appearance*. A high score on the appearance scale indicated that the user perceived this resource to be esthetically pleasing and thus easy to consume. The third

factor contained items focused on the overall ease of accessing this resource on the internet. Thus, this factor was labeled *accessibility*. A high score on the accessibility scale indicated that the user perceived this resource to be easy to find and without roadblocks to accessibility. These two factors emerged as hypothesized.

The fourth factor, systems support, was not hypothesized. These items focused on support needed from administration in order to implement the recommendations provided in the resource. A high score on the systems support scale indicated that the user would need more support from their system in order to implement the recommendations. However, the systems support factor also emerged in the URP-I and URP-A and thus is not entirely unexpected. Additionally, this factor contained two items that may warrant reverse coding if the resource is supposed to be used independently. Therefore, interpretability of this factor may be difficult.

All factors demonstrated acceptable levels of internal reliability as measured through Cronbach's alpha as hypothesized. Thus, the factors have been shown to measure the same construct through the EFA as well as through a secondary measure of reliability. In terms of social validity, the URP-WR exceeded the goal average score. Additionally, four of five categories met the goals. There was one category, systems support, that did not meet the goal. A higher score indicated that the user felt the need for support from their system to implement the assessment. This does not make the assessment unusable, just in need of support for use.

Generally, the URP-WR demonstrated an acceptable level of social validity as defined by the author. The goal scores were met for the scale as well as for four out of five categories measured by the URP-A. Thus, participants viewed the URP-WR as acceptable, understandable, feasible, and appropriate to system climate. They also indicated system support would be necessary to carry out the URP-WR. Thus, it may be difficult to implement independently.

Overall, this study demonstrated initial content validity of the URP-WR through the consensus-building task, a four-factor structure with 31 items through factor analysis, acceptable levels of internal validity within factors, and acceptable levels of social validity.

Implications for practice

This study began the development and validation of the URP-WR. In order to be used in practice, the URP-WR must undergo further evaluation and validation. Specifically, the implication inference must be tested in order to warrant use in a practical setting. Although it is not yet ready for immediate implementation, the initial development of the URP-WR is promising

because it provides a means of evaluating web-based resources that previously went unchecked. Education professionals will be able to use this tool to guide their decision making in order to make appropriate selections in their setting. Thus, the continuing development and improvement of it can result in an important and usable tool for education professionals.

Limitations

The first limitation of this study is the sample size. This study sought five participants per item, the ratio presented in Chafouleas et al. (2009). However, sample size recommendations for EFA vary in the available literature across several variable considerations. As factor analytic techniques began to emerge, some authors recommended sample sizes ranging from a 3:1 ratio to a 10:1 ratio of participants to items, and others recommended a minimum of 100 participants to conduct factor analysis (Cattell, 1978; Kline, 1994; Nunnally, 1967). More recently, a sample size "rule of thumb" has been elusive. de Winter, Dodou, and Wieringa (2009) showed that a recently developed "minimum" of 50 participants needed to interpret an exploratory factor analysis may be arbitrary as reliable results were found with fewer than 50 participants in simulated analyses with high levels of communality. Thus, the current sample size is small by many but not all criteria.

The second limitation is that a convenience sample was used. Thus, there may be characteristics of participants who elected to participate that differ from those who did not or were not selected. This is especially important to take into consideration concerning the social validity assessment, as those participants elected to continue after already completing 55 questions. Finally, some of the estimates of communality (h^2) fell below the suggested value of 0.50. Thus, the interpretation should be taken with caution until replication or future development shows more communality in those items.

Future directions

Future directions include further validation of the URP-WR. According to Kane's (2013) IUA model, the assumptions underlying an instrument include (a) scoring, (b) generalization, (c) extrapolation, and (d) implication. Further inferences, as well as expanding upon the scoring inference, need to be tested in the validation of the URP-WR. While testing other inferences, social validity should be reexamined. The social validity measure was used to evaluate the longer, 55-item draft version of the URP-WR. The length may have impacted user perceptions. The 31-item final URP-WR

should be evaluated for social validity again, and similarities or differences in results should be noted.

Additionally, future research should address possible other latent factors underlying web resource usability. Four factors were hypothesized based on an initial review of the literature, though this work illuminated additional factors that could be considered in future work. Future research can add additional considerations to the creation of quality educational web resources. For example, *coverage* was not considered as a distinct factor but rather was encompassed in credibility. Future work could consider the impact of explicit development and analyses of coverage items.

Finally, future research should address the connection between credibility and feasibility. This result was unexpected as credible, evidence-based interventions often run into roadblocks because of poor implementation fidelity. Poor implementation fidelity can emerge from feasibility issues. Therefore, the relationship found between items relating to feasibility and those relating to credibility should be further explored.

Conclusion

Given the increased emphasis on evidence-based practice in schools and the increased use of technology to access practice resources, an evaluation of the accessibility, appearance, plausibility, and systems support needed of web-based resources appears both timely and relevant. More objective evaluation of resources by these professionals should, in turn, support increased and improved implementation of evidence-based practices through the use of data-based decision making. A more informed, discerning consumption of web-based resources that are easily accessible, grounded in empirical evidence, and attend to feasibility of implementation has the potential to dramatically increase evidence-based practice use and subsequent outcomes for students. The creation of the URP-WR may also help bridge the research-to-practice gap by informing development of new web-based resources and revisions to existing web-based resources. Insights provided by consumers are critical to ensuring the wealth of available evidence-based practice is disseminated in a manner that is usable, attractive, and accessible for school psychologists and other educators. Furthermore, the use of the URP-WR may illuminate strengths and weaknesses of web-based resources. This provides web-based resource developers a manner of evaluating their own resources as they create them. The improvement in quality of resources coupled with the improvement in consumers' ability to evaluate resources will greatly improve the practice of web-based resource use.

Acknowledgement

The study reported in this manuscript was approved by the IRB of the authors' institution.

Disclosure statement

The authors whose names are listed immediately above certify that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge, or beliefs) in the subject matter or materials discussed in this manuscript.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

Benotsch, E. G., Kalichman, S., & Weinhardt, L. S. (2004). HIV-AIDS patients' evaluation of health information on the internet: The digital divide and vulnerability to fraudulent claims. *Journal of Consulting and Clinical Psychology*, *72*(6), 1004–1011. doi:10.1037/0022-006X.72.6.1004

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research*, *38*(1), 25–56. doi:10.1207/S15327906MBR3801_2

Buchanan, J., & Knock, N. (2001). Information overload: A decision making perspective. In M. Köksalan & S. Zionts (Eds.), *Multiple criteria decision making in the new millennium: Lecture notes in economics and mathematical systems* (p. 507). Springer.

Cattell, R. B. (1978). *The scientific use of factor analysis*. Plenum.

Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., & McCoach, D. B. (2009). Moving beyond assessment of treatment acceptability: An examination of the factor structure of the Usage Rating Profile – Intervention (URP-I). *School Psychology Quarterly*, *24*(1), 36–47. doi:10.1037/a0015146

Chafouleas, S. M., Miller, F. G., Briesch, A. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2012). *Usage rating profile – Assessment*. University of Connecticut.

Christ, T. J., Riley-Tillman, T. C., & Chafouleas, S. M. (2009). Foundation for the development and use of direct behavior rating (DBR) to assess and evaluate student behavior. *Assessment for Effective Intervention*, *34*(4), 201–213. doi:10.1177/1534508409340390

Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education*, *49*(6), 560–575. doi:10.1111/medu.12678

Cummings, J. A. (2011). Technology in the practice of school psychology: The future is past tense. In M. A. Bray & T. J. Kehle (Eds.), *Oxford library of psychology: The Oxford handbook of school psychology* (pp. 831–859). Oxford University Press.

de Winter, J. C., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, *44*(2), 147–181. doi:10.1080/00273170902794206

Fischer-Baum, R. (2017). What 'Tech World' did you grow up in? *The Washington Post*. Retrieved from https://www.washingtonpost.com/graphics/2017/entertainment/tech-generations/

Flaspohler, P. D., Anderson-Butcher, D., Paternite, C. E., Weist, M., & Wandersman, A. (2006). Community science and expanded school mental health: Bridging the research to practice gap to promote well-being and academic success. *Educational and Child Psychology*, *23*(1), 27. Retrieved from https://psycnet.apa.org/record/2006-06722-004

Greenwald, H. J., & O'Connell, S. M. (1970). Comparison of dichotomous and Likert formats. *Psychological Reports*, *27*(2), 481–482. doi:10.2466/pr0.1970.27.2.481

Greenwood, C. R., & Abbott, M. (2001). The research to practice gap in special education. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, *24*(4), 276–289. doi:10.1177/088840640102400403

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.

Hennessy, S., Rojas-Drummond, S., Higham, R., Marquez, A. M., Maine, F., Ríos, R.M., … Barrera, M. J. (2016). Developing a coding scheme for analyzing classroom dialogue across educational contexts. *Learning, Culture, and Social Interaction*, *9*, 16–44. doi:10.1016/j.lcsi.2015.12.001

Hintze, J. M., Volpe, R. J., & Shapiro, E. S. (2002). Best practices in the systematic direct observation of student behavior. *Best Practices in School Psychology*, *4*, 993–1006. https://gato-docs.its.txstate.edu/jcr:a137bf9a-b01a-4535-bf34-5c9eefd49986/Best%20Practices%20in%20the%20Systematic%20Direct%20Observation%20of%20Student%20Behavior%202002.pdf

Hunsley, J. (2007). Addressing key challenges in evidence-based practice in psychology. *Professional Psychology: Research and Practice*, *38*(2), 113–121. doi:10.1037/0735-7028.38.2.113

Isman, A., & Canan Gungoren, O. (2014). Digital citizenship. *Turkish Online Journal of Educational Technology-TOJET*, *13*(1), 73–77.

Johnson, A., Ledoux, M., Bains, B., Maggin, D., Buren, M., & Couch, L. (2019, February 28). *Special educator research utilization: Bridging the gap* [National Association of School Psychologists conference], Atlanta, GA.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. doi:10.1007/BF02291575

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. doi:10.1111/jedm.12000

Kline, P. (1994). *An easy guide to factor analysis*. Routledge.

Lydia M. Olson Library (2018). *Evaluating internet resources*. Retrieved from https://lib.nmu.edu/help/resource-guides/subject-guide/evaluating-internet-sources

McIntosh, K., Barnes, A., Eliason, B., & Morris, K. (2014). *Using discipline data within SWPBIS to identify and address disproportionality: A guide for school teams*. Technical Assistance Center on Positive Behavioral Interventions and Supports.

McNeill, J. (2019). Social validity and teachers' use of evidence-based practices for autism. *Journal of Autism and Developmental Disorders*, *49*(11), 4585–4594. doi:10.1007/s10803-019-04190-y

Miller, F. G., Neugebauer, S. R., Chafouleas, S. M., Briesch, A. M., & Riley-Tillman, T.C. (2013, August 5). *Examining innovation usage: Construct validation of the Usage Rating*

Profile-Assessment [poster session] of the American Psychological Association conference, Honolulu, HI.

Miller, M. (2012). 53% of organic search clicks go to first link. *Search Engine Watch*. Retrieved from https://www.searchenginewatch.com/2012/10/10/53-of-organic-search-clicks-go-tetrieved.fromho-first-link-study/

Morr, S., Shanti, N., Carrer, A., Kubeck, J., & Gerling, M. C. (2010). Quality of information concerning cervical disc herniation on the internet. *The Spine Journal: Official Journal of the North American Spine Society*, *10*(4), 350–354. doi:10.1016/j.spinee.2010.02.009

Noell, G. H., & Gansle, K. A. (2014). Research examining the relationships between consultation procedures, treatment integrity, and outcomes. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation* (2nd ed., pp. 386–408). Taylor & Francis Group/Routledge.

Nunnally, J. C. (1967). *Psychometric theory*. McGraw-Hill.

Pinkerton, J. A. (2000). *WebCrawler: Finding what people want* [Unpublished doctoral dissertation]. University of Washington.

Raines, J. C., & Raines, J. C. (2008). *Evidence based practice in school mental health*. Oxford University Press.

Ream, E., Blows, E., Scanlon, K., & Richardson, A. (2009). An investigation of the quality of breast cancer information provided on the internet by voluntary organisations in Great Britain. *Patient Education and Counseling*, *76*(1), 10–15. doi:10.1016/j.pec.2008.11.019

Ribble, M. (2012). Digital citizenship for educational change. *Kappa Delta Pi Record*, *48*(4), 148–151. doi:10.1080/00228958.2012.734015

Ribble, M. S., & Bailey, G. D. (2004). Digital citizenship focus questions for implementation. *Learning & Leading with Technology*, *32*(2), 12–15. https://eric.ed.gov/?id=EJ695806

Rosenfield, S. (2000). Crafting usable knowledge. *The American Psychologist*, *55*(11), 1347–1355. doi:10.1037/0003-066X.55.11.1347

Schrock, K. (2019). *Schrock guide.* Retrieved from https://www.schrockguide.net/

Scott, T. M., Hirn, R. G., & Barber, H. (2012). Affecting disproportional outcomes by ethnicity and grade level: Using discipline data to guide practice in high school. *Preventing School Failure: Alternative Education for Children and Youth*, *56*(2), 110–120. doi:10.1080/1045988X.2011.592168

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*(2), 203–214. doi:10.1901/jaba.1978.11-203