

Journal Pre-proof

The Design and Methodology for a Pilot Study of Home and Community-Based Services Outcome Measures

Alec Nyce, Matt Roberts, Renáta Tichá, Brian Abery



PII: S1936-6574(24)00059-1

DOI: <https://doi.org/10.1016/j.dhjo.2024.101628>

Reference: DHJO 101628

To appear in: *Disability and Health Journal*

Received Date: 3 November 2023

Revised Date: 29 March 2024

Accepted Date: 12 April 2024

Please cite this article as: Nyce A, Roberts M, Tichá R, Abery B, The Design and Methodology for a Pilot Study of Home and Community-Based Services Outcome Measures, *Disability and Health Journal*, <https://doi.org/10.1016/j.dhjo.2024.101628>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier Inc.

The Design and Methodology for a Pilot Study of Home and Community-Based Services Outcome Measures

Alec Nyce^a M.S. M.A., Matt Roberts^a M.A., Renáta Tichá^a Ph.D., Brian Abery^a Ph.D.

^aInstitute on Community Integration, University of Minnesota

Keywords: HCBS outcome measurement, psychometrics, person-centered, intellectual and developmental disabilities, HCBS disability populations

Abstract word count: 246

Manuscript word count: 4321

Number of references: 33

Number of figures/tables: 3

Author Note

Alec Nyce M.S. M.A.  <https://orcid.org/0000-0003-0607-8522>

Matt Roberts M.A.  <https://orcid.org/0000-0003-4064-8509>

Renáta Tichá Ph.D.  <https://orcid.org/0000-0001-6431-5261>

Brian Abery Ph.D.  <https://orcid.org/0000-0003-1998-6511>

The measures being piloted in this study were developed at the authors' institute of employment for endorsement of use at the national level but will be provided free-of-charge once development is concluded. Otherwise, we have no known conflict of interest to disclose. This publication was supported by federal grant/cooperative agreement #90RT5039 from the National Institute on Disability, Independent Living, and

Rehabilitation Research (NIDILRR) within the Administration for Community Living (ACL).

Correspondence concerning this article should be addressed to Alec Nyce, Institute on Community Integration, University of Minnesota, 2025 East River Parkway, Minneapolis, MN 55414. Phone: (612) 625-5381. Email: nycex001@umn.edu

Acknowledgments

The authors would like to thank all the participants and HCBS providers for sharing their time and energy with us. We would also like to thank our partners at the Temple University Collaborative in Community Inclusion and all the members of the RTC/OM team over the years who helped with data collection.

Author Contributions

All authors made substantial contributions to the conception and idea of the manuscript. Nyce was the primary contributor to the data analysis as well as writing the results section and a majority of revisions. Roberts and Nyce made substantial contributions to the methods section. Roberts, Tichá, and Abery made substantial contributions to the context of for the article in the introduction, as well as providing background information on the RTC/OM and HCBS measurement. All authors made substantial contributions to the discussion section by interpreting results and providing context.

1
2
3
4
5
6
7
8

**The Design and Methodology for a Pilot Study of Home and Community-Based
Services Outcome Measures**

Journal Pre-proof

9

Abstract

10 **Background.** The Research and Training Center on HCBS Outcome Measurement
11 (RTC/OM) developed and piloted measures in six domains to assess the outcomes
12 experienced by HCBS recipients. These measures were based upon the revised
13 National Quality Forum's HCBS Outcome Measurement framework.

14 **Objective.** The background and rationale for the pilot study are outlined along with the
15 research design, sampling frame, and psychometric and statistical methods used. In
16 addition, administration feasibility for all measures are described. Finally, a summary of
17 results across all measures is provided. Detailed results for individual outcome measure
18 domains are left to forthcoming publications.

19 **Methods.** Measure construct under study were piloted on a sample of 107 participants
20 identified as receiving HCBS or HCBS-like services and experiencing one of five
21 disabilities: intellectual and developmental, age-related, or physical disabilities as well
22 as Traumatic/Acquired Brain Injury and Serious Mental Health Conditions. Participants
23 were interviewed either in-person or through HIPAA compliant online video
24 conferencing over one to two sessions. Psychometric evidence was evaluated with
25 internal consistency and test-retest reliability, as well as inter-observer agreement.
26 Nonparametric methods were used to test for group comparisons.

27 **Results.** Initial reliability and validity results of outcomes on five measures were good to
28 excellent. No significant group differences between disability groups were found.

29 **Conclusions.** The psychometric evidence for the tested measures is very promising.
30 Only two of the six measures required significant changes prior to their use in an
31 upcoming field study. Details on results and revisions for individual measures will
32 appear in later publications.

33

Introduction

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

High-quality, psychometrically-sound measurement is essential to obtaining useful information about the impact of Home- and Community-Based Services (HCBS) on recipients. HCBS outcome measures are psychometrically sound when they have evidence to support the argument that data obtained from them are both reliable and valid¹. Without sufficient evidence supporting the reliability and validity of measure outcomes, decisions for which they are used may not be adequately informed. In the absence of reliable, valid outcomes, federal and state agencies cannot make accurate judgements with respect to compliance with regulations. States are unable to make informed decisions about appropriate funding allocations, and provider agencies are left in the dark when addressing areas in need of quality improvement. At the individual level, HCBS recipients are left with inaccurate and potentially misleading information when it comes to making decisions about which provider(s) or service(s) to select.

The National Core Indicators In-Person Survey² (NCI IPS), the Council for Quality and Leadership (CQL) Personal Outcome Measures³ (POMs), and Consumer Assessment of Healthcare Providers and Systems Home and Community-Based Services⁴ (CAHPS) are three popular approaches used to assess HCBS outcomes. The NCI IPS was built for state level assessment and monitoring and is useful for monitoring quality assurance. The CQL POMs is a well-developed and validated tool with good psychometric properties while also being relevant to a wider group of disability populations. Both the CQL POMs and the HCBS CAHPS have the advantage of being developed for use at the individual and provider level⁵.

A shared limitation for many existing outcome measures is that many items cannot be considered person-centered⁶ (i.e., responded to by the person with a

57 disability and eliciting the extent to which their individual need or desired level of an
58 outcome is being met. See Roberts & Abery⁵ for a nuanced discussion of person-
59 centeredness.) Another shared limitation is the lack of evidence for use longitudinally or
60 their sensitivity to measuring change⁵. In fact, the Human Services Research Institute
61 which administers the NCI program did not develop the measures to be used
62 longitudinally and explicitly state that the measures are not to be used in this manner. In
63 addition, states have only employed cross-sectional sampling methods when
64 administering the NCI in the past. Consequently, there is no empirical evidence
65 supporting its longitudinal use. Moreover, the NCI is limited by being developed
66 explicitly for state-level data and is not applicable for use at the individual or provider
67 level.

68 The HCBS CAHPS, while ostensibly developed for use at the individual and
69 provider level, lacks sufficient psychometric evidence to justify its use⁷. There are
70 several results that were reported in the HCBS CAHPS final report⁴. Reliability
71 coefficients fell well below recommendations of scientific acceptability, exploratory factor
72 analyses were reported as confirmatory, factor analysis results (e.g., factor loadings)
73 were unreported, and no external validity testing was performed despite claims to the
74 contrary. In addition, item removal and revisions were performed after psychometric
75 testing had concluded, so the final recommended version of the measure had no data
76 on outcome reliability and validity.

77 The Rehabilitation Research and Training Center on HCBS Outcome
78 Measurement (RTC/OM) funded by the National Institute on Disability, Independent
79 Living, and Rehabilitation Research (NIDILRR) under the Administration on Community
80 Living (ACL) was tasked with developing person-centered HCBS quality and outcome

81 measures based on the National Quality Forum's (NQF) HCBS Framework⁸. To
82 accomplish this goal, the RTC/OM first validated the importance of NQF framework
83 domains and subdomains with a national sample of HCBS recipients, family members,
84 providers, and program administrators/policy makers using a participatory planning and
85 decision making (PPDM) process⁶. This process gathered participant input on the
86 importance of each NQF domain and subdomain. RTC/OM staff subsequently used
87 these data in conjunction with a gap analysis to revise the NQF framework. The two
88 changes to the framework were elevating the concept of transportation to a subdomain
89 with the community inclusion domain, and employment from being a subdomain within
90 the community inclusion domain to being its own domain. Subsequently, technical
91 expert panels composed of content experts, providers, family members and people with
92 disabilities prioritized the HCBS quality and outcome measures for development. The
93 six measures produced from this effort focus on the following outcomes: (1) Meaningful
94 Activity, (2) Social Connectedness, (3) Choice and Control, (4) Employment, (5)
95 Transportation, and (6) Freedom from Abuse and Neglect (see Table 1). The measures
96 were also designed to have features lacking in the previously-discussed measures (i.e.,
97 ability to be used longitudinally, modular in format, and free to use.)

98 **(INSERT TABLE 1 ABOUT HERE)**

99 For each measure, the team developed blueprints summarizing background
100 literature for each construct and outlining the structure of each concept. To develop
101 specific items, the team first drafted guiding questions for each measure. Guiding
102 questions specify the intended inferences and supporting assumptions for each
103 measure (i.e., questions the measure is intended to answer). These include specific
104 inferences related to each measure domain and subdomain. Items from existing HCBS

105 instruments were reviewed to ensure full coverage of the domains/subdomains. Based
106 on this review, the RTC/OM team created initial items to capture the domains and
107 subdomains identified in the blueprint for each measure concept. Following the
108 completion of draft items, a panel of content and measurement experts provided ratings
109 and qualitative feedback as to whether each item measured its intended domain.
110 Measures and conceptual definitions were revised following the expert panel review. An
111 adapted cognitive testing protocol^{9,10} based on the Cognitive Aspects of Survey
112 Methodology (CASM) model^{11,12} was also completed with five members of each target
113 population for each measure and the data were used for final measure revisions prior to
114 the pilot study. A recent technical report¹³ has more information on the expert panel and
115 cognitive testing protocols and results.

116 Each measure was conceptualized as having two tiers. Tier 1 (i.e., *global* items)
117 consists of a small number of questions assessing broad aspects of the construct.
118 These items are intended to gain a general sense of participants' impressions of the
119 concept. Tier 2 (i.e., *specific* items) contains granular items that provide more detailed
120 and, in some cases, actionable information. Measures were constructed so that multiple
121 specific questions in Tier 2 were "clustered" with a global question in Tier 1. Items in
122 these clusters were hypothesized to have stronger relationships with each other
123 (quantified by linear correlation) than with items outside these clusters. Broadly, these
124 two-tiered and clustering approaches were utilized to: 1) test the validity of the
125 constructs, 2) gather insight with respect to how specific items are related to the broad
126 measure construct, and 3) potentially reduce the length of measures¹³.

127 The item responses to the RTC/OM measures were predominantly structured to
128 be *scalable*. Scalable items have response formats that are ordered. They define the

129 underlying construct and are collectively used as multiple indicators to “sample” a
130 person’s level on the construct. The benefit of this approach is that combining item
131 responses (e.g., sum scores) has a meaningful interpretation. The three scalable item
132 formats used for the measures were a four-point, bi-polar Agreement scale (Strongly
133 Disagree, Disagree, Agree, Strongly Agree) and two four-point, uni-polar Frequency
134 scales (Never, Sometimes, Most of the Time, Always; None, Rarely, Some of the Time,
135 Most of the Time). When items were not scalable (e.g., demographic) they were either
136 rekeyed into a composite variable or excluded from psychometric analysis.

137 The measures were constructed as primarily positively-worded items (e.g., “I
138 participate in activities that are meaningful to me”). There are downsides to this
139 approach. Developing measures with only positively-worded items can mask other
140 problematic response styles, notably acquiescence responding (always responding
141 “positively” regardless of content). If participants engage in acquiescence responding it
142 is difficult to distinguish them from participants whose positive responses are genuine or
143 control for these response styles with methods such as factor analysis¹⁴. A balance of
144 positively- and negatively-worded items would then seem ideal, but previous research
145 has shown that negatively-worded items can cause confusion and lead to errant
146 responding, especially for those with intellectual or developmental disabilities¹⁵⁻¹⁸.
147 Moreover, researchers have suggested that negatively-worded items are themselves
148 substantively distinct constructs (not merely antonyms) or give rise to extraneous
149 “method factors”, the latter being related to the reading ability of responders¹⁹p.113-115.

150

151 **Methods**

152 **Pilot Study Procedure**

153 Approval for this study was obtained by the Institutional Review Boards (IRBs) of
154 the collaborating universities. All participants either provided consent themselves, or
155 when not their own legal guardian, assent alongside the consent of a legally authorized
156 representative. Participants were reminded in the measure scripting that they could
157 withdraw at any time or pass/refrain from answering any question during the interview.

158 Measures were administered using a structured, but conversational, interview
159 format. Two modes of administration were tested during the pilot study: in-person
160 ($n=85$) and videoconferencing ($n=22$). In-person interviews entailed completing the
161 interview with the person in a private space, such as a meeting room. Videoconference
162 administration used a HIPAA compliant version of the Zoom meeting platform to
163 administer the measures remotely. All interviewers were trained in consent procedures
164 and research ethics, disability etiquette, interviewing procedures for both modes of
165 administration, and in use of the data collection software.

166 Measure protocols were identical between the two modes of administration
167 except for response option cards being displayed by the interviewer via screen sharing
168 in the videoconference interviews and a physical card during in-person interviews. Data
169 were collected by trained interviewers using the Qualtrics survey platform Offline App.
170 For items with an agreement scale, one of two visual scales was presented to aid
171 responding. All participants with even-numbered identification (ID) numbers were
172 offered a text-only version of the agreement scale. Participants with odd-numbered IDs
173 were offered the same text agreement scale with the addition of emojis above the text.

174 This procedure was used to test whether the added emoji scale aided responding for
175 some individuals.

176 On a portion of interviews, a second interviewer collected agreement, fidelity, and
177 etiquette of measure administration data. *Interrater agreement* was assessed through
178 having the second interviewer input participant's responses in a duplicate survey.
179 *Fidelity* and *etiquette* were assessed by a second interviewer answering a set of
180 questions to document and ensure that primary interviewers were following
181 administration protocols.

182 *Test-retest reliability* (i.e., the coefficient of stability^{20(p.117)}) was evaluated using a
183 retest target sample of five participants from each target disability group. The attempted
184 retest timeframe was 10-14 days. This timeframe was chosen to minimize rehearsal
185 effects and decrease the likelihood participants would experience true change in the
186 construct²¹. A similar time frame has been used by developers of other social inclusion
187 and quality of life measures²²⁻²⁴.

188 **Sample**

189 The pilot sample was recruited through local HCBS providers and included
190 persons who were 18 years of age or older, currently receiving HCBS or HCBS-like
191 services and residing in an upper-Midwestern and Mid-Atlantic state. HCBS-like
192 services are those that resemble services covered by federal HCBS waivers but are not
193 directly funded through a waiver. All pilot participants had received diagnoses of
194 experiencing age-related, physical, intellectual/developmental, or psychiatric disability,
195 or traumatic/acquired brain injury and were required to pass a capacity-to-consent
196 assessment (UBACC²⁵, adapted). Efforts were made during recruitment to ensure
197 diverse representation within each disability group including age, gender, disability type,

198 and intensity-of-support needed or functional level. People with primarily sensory
199 disabilities and participants who could not communicate their own responses were
200 excluded from the sample.

201 A total of 107 people were recruited for the study with 60 from the upper Midwest
202 and 47 from the Mid-Atlantic state. Measures were administered in a randomized order
203 during interviews. Some participants did not complete all measures during the first
204 interview and subsequently completed them in a follow-up session. In addition, the
205 *Abuse and Neglect* measure was not taken by two participants due to concerns with the
206 measure content. The final sample sizes for measure domains assessed were: *Abuse*
207 *and Neglect* (N=100), *Choice and Control* (N=100), *Employment* (N=101), *Meaningful*
208 *Activity* (N=101), *Social Connectedness* (N=105), and *Transportation* (N=102). Fifteen
209 individuals who initially expressed interest did not participate in the study. Of these,
210 seven people did not meet the inclusion criteria, three did not pass the capacity to
211 consent assessment, two declined to consent, two withdrew after consenting; and one
212 was unreachable following consent.

213 Participant demographics are provided in Table 2. A majority of the sample self-
214 identified as White ($n=72$), Black/African American ($n=18$), or as more than one race
215 ($n=6$). Fewer than five participants identified as American Indian/Native Alaskan,
216 Hispanic/Latino, Asian, or Native Hawaiian/Pacific Islander. Two participants did not
217 know or refused to answer. Compared to U.S. Census³³ state prevalence data, our
218 sample overrepresented Black/African American individuals (7-10%) and
219 underrepresented Hispanic/Latino individuals (6-8%).

220 Fewer participants had an age-related disability (ARD) or a traumatic brain injury
221 (TBI) compared with the other disability groups. The ARD population was much older

222 than the other four disability groups. Twenty-nine percent of participants ($n = 31$, 29%)
223 resided in rural areas, with most living in their own home ($n = 49$) with other people with
224 disabilities and support staff ($n = 32$). The primary source of support for participants was
225 from paid staff ($n = 51$) or unpaid family members ($n = 26$), with participants receiving
226 support from paid or unpaid family much more likely to live with family.

227

228

(INSERT TABLE 2 ABOUT HERE)

229 **Feasibility Analysis**

230 One of the primary purposes of the study was to evaluate the feasibility of the
231 measures. Feasibility in this context means the measures are easy to administer and
232 score, need minimal interviewer training, can be completed in a timely fashion (i.e., ten
233 to fifteen minutes per measure), and are relevant to the participants being assessed.
234 Feasible measures are relevant both when the item content is of consequence to
235 participants and when response options accurately capture the range of participant
236 experiences. This feasibility review also included feedback from both interviewers and
237 participants with disabilities about the interview process. Alongside psychometric
238 analysis, feasibility results were used to revise items.

239 **Administration Procedures**

240 Individual interviewers provided notes and feedback about the interviewing
241 process and how participants responded to the content and structure of the measures.
242 The impressions of the interviewers were generally positive with administrative
243 concerns noted and used to improve the interview process. Interviewer notes also
244 included feedback on participant understanding of items and suggestions made by
245 participants with respect to how items could be improved. Qualitative feedback was

246 summarized and taken into consideration following study completion to make item
247 wording revisions prior to the field study. For example, on the Transportation measure
248 additional barriers to using preferred transportation were added. In addition, on items
249 assessing whether a participant receive sufficient support, a few individuals reported not
250 needing support in that area. To address this, a “I don’t need support in this area”
251 response option was added for these items.

252 The measures were administered using a video-conference (VC) mode to a small
253 sample of participants ($n=22$) to obtain preliminary evidence regarding the feasibility of
254 this approach. This VC sample was different from the whole sample in three ways: 20
255 (91%) VC participants were from non-rural areas (compared with 71% of the full
256 sample); 10 (45%) experienced age-related disabilities (16% in the full sample); and 19
257 (86%) were from the Mid-Atlantic state (44% in the full sample). In addition, one
258 interviewer on the research team did most VC interviews but they did not indicate major
259 problems using this mode of administration. Minor technical difficulties were reported for
260 two of the interviews (i.e., slow internet connection, the Zoom application not working).
261 Despite the occasional minor difficulties, data collection was completed for all six
262 measures with each participant administered via this mode. Feedback data indicated a
263 generally positive experience that did not differ substantially from participants
264 interviewed using the in-person mode.

265 ***Measure completion times***

266 Individual completion times (in minutes) for each measure were provided by the
267 Qualtrics survey tool. Qualtrics tracks measure time from when the interviewer opened
268 the survey to when it is submitted. Outliers were removed prior to calculating measure
269 completion times. These outliers were defined as a measure completion time greater

270 than 60 minutes. The number of outliers for each measure ranged from 2 to 5. The
271 values of most outliers were greater than one thousand minutes (e.g., unintentionally
272 opening of the survey link or the submission timed out from closing the survey without
273 manually submitting the responses). Mean completion times across measures ranged
274 from M=10.2 (SD=7.3; Mdn=7.9; *Social Connectedness*) to M=18.2 (SD=8.6; Mdn=15.4;
275 *Choice and Control*). Distributions of response times for each measure were right-
276 skewed.

277 **Psychometric and Statistical Analysis**

278 Only data collected during a participant's first interview were used to compute
279 statistical results, unless otherwise noted (e.g., retest data). Data across in-person and
280 video-conference interviews were combined since there were no overt systematic
281 differences in responses between these two groups. Data analyses were performed in
282 R (version 4.0.0 or later). For all items and demographic variables, univariate
283 descriptive statistics (e.g., mean, range, median, standard deviation) were calculated
284 using the `psych` package²⁶. For scalable items, Pearson product moment (PPM)
285 correlations, internal consistency reliability, and test-retest correlations are reported.
286 Missing data for univariate statistics were ignored, whereas full information maximum
287 likelihood²⁷p.86 (FIML) was used to estimate PPM correlations. The latter approach
288 sometimes led to improper correlations matrices²⁸ which were corrected using matrix
289 smoothing techniques in the `lavaan` package²⁹.

290 *Internal consistency reliability* of measure responses was assessed with
291 Cronbach's alpha, a well-researched and widely-used method that was listed by the
292 NQF as a method to demonstrate scientific acceptability³⁰p.55. Internal consistency was
293 also used to evaluate the two-tiered and clustering approaches. *Test-retest reliability*

294 was evaluated using PPM correlations between participant total scores across time
295 points, where total scores were created by summing scalable items on each measure.
296 *Inter-observer agreement* was assessed at the measure level using the average (%)
297 agreement across all pairs of interviewers and participants.

298 Differences in responding between measure response scales (emoji vs text) and
299 disability populations were assessed across all measures. Traditionally, methods such
300 as the *t*-test and ANOVA have been used to perform such comparisons³¹. However,
301 distributions of item responses and measure sum scores did not meet the assumptions
302 of normality and equal variance. Due to this, nonparametric statistical tests were used
303 instead. The *Wilcoxon signed rank test*^{32|p.56} (WSRT) and *Kruskal-Wallis ANOVA*^{32|p.204}
304 were used in lieu of the paired-sample *t*-test and classic ANOVA, respectively. These
305 analyses were not included in the original study proposal, so sample sizes were not
306 determined by prospective power analyses.

307

Results

308 Overview of Results

309

310

311

312

313

314

315

316

A summary of internal consistency reliability, test-retest reliability, and interobserver agreement results are provided in Table 3. Estimates include all scalable items unless otherwise noted. In each cell of internal consistency results, '*Global*' indicates estimates using only Tier 1 items. The one exception is the *Employment* measure, which had separate items for participants who were either employed (*Job Experiences* [JE]) or unemployed and looking for work (*Barriers* [B]). Estimates were also obtained for two *Choice & Control* subscales: *Services & Supports* (S&S) and *Personal Choices & Goals* (PC&G).

317

318

319

320

321

322

323

324

325

326

327

328

Five measures had good to excellent internal consistency reliability ($\alpha \geq .80$). The *Barriers*, *Support & Services*, and *Personal Choices & Goals* subdomains would meet the criteria for basic research ($\alpha \geq .70$). The *Abuse and Neglect* measure was not adequate ($\alpha = .62$). When considering only global items, only two measures met the minimum standard for basic research ($\alpha \geq .70$). We also found that all measures and subscales had adequate to excellent test-retest reliability ($\geq .70$) and excellent interobserver agreement ($\geq .89$). Finally, no significant differences were found between distributions of responses between disability groups nor between participants receiving different response scales (i.e., emoji vs text). A detailed presentation of these results (e.g. effect sizes, confidence intervals) for each measure will be in forthcoming manuscripts (in preparation).

329

Discussion

330

331

332

333

334

335

336

The RTC/OM developed and piloted measures assessing the domains of *Meaningful Activity, Social Connectedness, Choice and Control, Employment, Transportation, and Freedom from Abuse and Neglect*. Results from the pilot study led to minor changes to some measures (*Meaningful Activity, Social Connectedness*), while more substantial revisions were made to others (*Choice and Control, Abuse and Neglect, Employment*). Details of changes to individual measures are left to their respective articles within this series.

337

(INSERT TABLE 3 ABOUT HERE)

338

339

340

341

342

343

344

345

346

347

348

349

Overall, results indicate that the approach RTC/OM took to developing these measure concepts has led to outcome measures with sound psychometric properties that are feasible to implement. There were no significant differences found between disability groups on their reported outcomes within each measure. This is an important finding given the recent focus of CMS to have measures available that cut across multiple disability populations. We also found minimal differences between participant experiences when being interviewed either in-person or remotely via videoconferencing which suggests the usefulness of the measures who are difficult to interview in-person (e.g., those in rural settings). Conclusions about the feasibility of the measures and psychometric estimates did not differ based on the environments in which the data were collected; and participants' paired test-retest reliability interviews were all conducted with the same restrictions in place (unrestricted or restricted).

350

351

352

At this time we do not have sufficient data to support the two-tiered approach to measurement. When considering global items, only two measures met minimum standards. In addition, the hypothesis that intercorrelations among items within item

353 clusters would be greater than their intercorrelations with items outside the cluster
354 (convergent and discriminant validity) was not supported. However, this analysis
355 approach was limited by the study's sample size, which forwent the use of factor
356 analysis and other data-reduction techniques. The two-tiered conceptualization will be
357 assessed with factor analysis and other techniques in a national field study currently
358 underway.

359 This study provides preliminary evidence for the use of newly developed
360 outcomes measure specifically to be used at the provider level. Given their modular
361 nature (i.e., each measure can be used on its own) the measures have practical utility
362 as providers can select measures specific to outcomes on which they desire to focus or
363 potentially for particular clients, thus promoting a person-centered approach and time
364 efficiency. The domains/subdomains the measures cover have been selected to include
365 key areas of service provision in both residential and employment settings. This new set
366 of measures have the potential to be used as an alternative to existing measures and
367 instruments with psychometric data back up their quality. There is also potential for the
368 measures to be used at the HCBS program or state policy level. The evidence for the
369 validity of inferences at these levels of implementation would first need to be
370 established with data that represents the context of application.

371 **Limitations**

372 There were some limitations to the pilot study that may have affected the
373 representativeness of the sample and results. First, the requirement to successfully
374 pass an informed consent assessment (UBACC²⁵) limited the sample to those HCBS
375 recipients with less severe communication and cognitive disabilities. In the national field
376 study of the measures, versions of the measures designed for a proxy to answer will be

377 tested allowing for responses to be obtained from someone who knows the person well.
378 In addition, both the items and the interview process require English proficiency and no
379 psychometric data are available to address measurement across different cultural
380 backgrounds. It is plausible that constructs such as Social Connectedness will function
381 differently for those from different backgrounds and a diverse sample of individuals is
382 needed to explore this further. Another limitation in the study was the impact of the
383 COVID-19 pandemic. Pandemic-related restrictions began midway through data
384 collection for the pilot study. In the Mid-Atlantic state, this resulted in 20 participants
385 having data collected during a stay-at-home order. However, the purpose of the study
386 was to test and determine the feasibility and preliminary estimates of psychometric
387 characteristics of the measures and not to assess actual levels of an outcome at a
388 particular time and place. Moreover, the video-conference interviews completed during
389 this time did not result in feasibility concerns.

390 **Implications and Future Directions**

391 This article provides preliminary evidence regarding the development and
392 psychometric qualities of the new RTC/OM measures. For the purposes of this paper,
393 the revised measures tested during the RTC/OM's national field study will be subject to
394 a wider breadth of psychometric and statistical analyses to ensure their compliance with
395 general psychometric criteria in the field and criteria for an endorsement process at a
396 national level. An additional cross-validation approach to exploratory and confirmatory
397 factor analysis will be performed to test for and finalize the measure structure and item
398 content. Differential item functioning across all measures will ensure that items are not
399 biased against, or performing differently for, different disability groups. Finally, data are
400 being collected for participants across three time points which allows for a longitudinal

401 approach, which has additional benefits. The longitudinal psychometric properties of the
402 measure can be assessed, such as the measure's sensitivity to change and longitudinal
403 invariance. In addition, a mixed-effects modelling approach can be used to assess
404 different predictors of changes on the measure outcomes. This work is currently
405 ongoing with the intention to publish future articles that address these different areas.

Journal Pre-proof

406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422

Conclusion

For decades, the HCBS measurement field has been evolving toward using measurement tools that are person-centered. The RTC/OM measures are being developed with the intent of providing tools that can be used at the *provider* level to determine HCBS quality and outcomes. The RTC/OM developed the measures with the intention of supplementing other measurement efforts in the field with additional person-centered tools for provider organizations to use for quality improvement and organizational decision making within the selected outcome domains. Following the completion of the field study, the measures will undergo final revisions and the RTC/OM will provide free technical assistance and training to support their use across the country in a responsible and intended manner. The training will be designed for HCBS provider organizations or other entities interested in using the measures to monitor progress resulting from quality improvement efforts or person-centered services on the outcomes of HCBS recipients within particular domains or subdomains. Further testing will be needed as to whether evidence supports their use for other applications such as provider score cards or compliance with federal statutes such as the HCBS Final Rule.

423

References424 1 AERA, APA, & NCME. (2014). *Standards for Educational and Psychological Testing.*

425 American Educational Research Association.

426 2 National Association of State Directors of Developmental Disabilities Services

427 (NASDDDS) and Human Services Research Institute (HSRI) (2021). National

428 Core Indicators: NCI-IDD indicators. Retrieved from

429 <https://idd.nationalcoreindicators.org/wp->430 [content/uploads/2022/09/NCI_IDD_Indicators_FINAL_21-22.pdf](https://idd.nationalcoreindicators.org/wp-content/uploads/2022/09/NCI_IDD_Indicators_FINAL_21-22.pdf)431 3 The Council on Quality and Leadership (CQL) (2017). *Personal Outcome Measures:*432 *Measuring personal quality of life* (3rd ed.) Towson, Maryland.433 4 Truven Health Analytics (THA) (2017). *The development, testing and credentialing of*434 *the CAHPS® Home and Community Based Services Survey.*

435 5 Roberts, M. A. & Abery, B. H. (2023). A person-centered approach to home and

436 community-based services outcome measurement. *Frontiers in Rehabilitation*437 *Sciences*, 4(2023). <https://doi.org/10.3389/fresc.2023.1056530>

438 6 Rehabilitation Research and Training Center on Outcome Measurement. (RTC/OM).

439 (2017). *Stakeholder Input: Identifying Critical Domains and Subdomains of HCBS*440 *Outcomes.* Retrieved from <https://rtcom.umn.edu/phases/phase-1-stakeholder->441 [input](https://rtcom.umn.edu/phases/phase-1-stakeholder-input)

442 7 Nyce, A., Abery, B., & Tichá, R. (2020). An analysis of the HCBS CAHPS survey

443 [Unpublished manuscript]. Institute on Community Integration, University of

444 Minnesota.

445 8 National Quality Forum (NQF). (2016). *Quality in Home and Services to Support*446 *Community Living: Addressing Gaps in Performance Measurement.* Washington,

- 447 DC. Retrieved from
448 [http://www.qualityforum.org/Publications/2016/09/Quality in Home and Commu-](http://www.qualityforum.org/Publications/2016/09/Quality_in_Home_and_Community-)
449 [nity-](http://www.qualityforum.org/Publications/2016/09/Quality_in_Home_and_Community-)
450 [Based Services to Support Community Living Addressing Gaps in Perform-](http://www.qualityforum.org/Publications/2016/09/Quality_in_Home_and_Community-)
451 [ance Measurement.aspx](http://www.qualityforum.org/Publications/2016/09/Quality_in_Home_and_Community-)
- 452 9 Willis, G. B. (2005). *Cognitive interviewing : a tool for improving questionnaire design*.
453 Sage Publications.
- 454 10 Beatty, P. C., & Willis, G. B. (2007). *The Practice of Cognitive Interviewing*. *The*
455 *Public Opinion Quarterly*. Oxford University Press American Association for
456 Public Opinion Research. <https://doi.org/10.2307/4500375>
- 457 11 Tourangeau, R. (2018). The survey response process from a cognitive viewpoint.
458 *Quality Assurance in Education*, 26(2), 169–181. [https://doi.org/10.1108/QAE-06-](https://doi.org/10.1108/QAE-06-2017-0034)
459 [2017-0034](https://doi.org/10.1108/QAE-06-2017-0034)
- 460 12 Tourangeau, R. (1984), “Cognitive science and survey methods”, in Jabine, T., Straf,
461 M.L., Tanur, J.M. and Tourangeau, R. (Eds), *Cognitive Aspects of Survey*
462 *Design: Building a Bridge Between Disciplines*, National Academy Press,
463 Washington, DC, pp. 73-100.
- 464 13 Research and Training Center on HCBS Outcome Measurement (RTC/OM) (2021).
465 RTC/OM Measure Development Blueprint [Unpublished Manuscript].
- 466 14 D'Urso, E. D., Tijmstra, J., Vermunt, J. K., & De Roover, K. (2023). Awareness is
467 bliss: How acquiescence affects exploratory factor analysis. *Educational and*
468 *Psychological Measurement*, 83(3), 433-472.
- 469 15 Kramer, J. M., & Schwartz, A. (2017). Reducing Barriers to Patient-Reported
470 Outcome Measures for People With Cognitive Impairments *Archives of Physical*

- 471 Medicine and Rehabilitation. Archives of Physical Medicine and Rehabilitation,
472 98, 1705–1720. <https://doi.org/10.1016/j.apmr.2017.03.011>
- 473 16 Beddow, P., Elliott, S., & Kettler, R. (2013). Test Accessibility: Item Reviews and
474 Lessons Learned from Four State Assessments. Education Research
475 International, 2013, 1-12.
- 476 17 Weems, G. H., Anthony J. Onwuegbuzie & Kathleen M.T. Collins (2006) The Role of
477 Reading Comprehension in Responses to Positively and Negatively Worded
478 Items on Rating Scales, Evaluation & Research in Education, 19:1, 3-20, DOI:
479 10.1080/09500790608668322
- 480 18 Finlay, W. M. L., & Lyons, E. (2001). Methodological issues in interviewing and using
481 self-report questionnaires with people with mental retardation. Psychological
482 Assessment, 13(3), 319–335. [https://doi-org.ezp1.lib.umn.edu/10.1037/1040-](https://doi-org.ezp1.lib.umn.edu/10.1037/1040-3590.13.3.319)
483 [3590.13.3.319](https://doi-org.ezp1.lib.umn.edu/10.1037/1040-3590.13.3.319)
- 484 19 Bandalos, D. L. (2018). Measurement theory and applications for the social
485 sciences. New York, NY: The Guilford Press.
- 486 20 Crocker, L. & Algina, J. (2008). Introduction to classical and modern test theory.
487 Mason, OH: Cengage Learning.
- 488 21 Polit, D. F. (2014). Getting serious about test–retest reliability: a critique of retest
489 research and some recommendations. Quality of Life Research, 23(6), 1713-
490 1720.
- 491 22 Marino-Francis, F., & Worrall-Davies, A. (2010). Development and validation of a
492 social inclusion questionnaire to evaluate the impact of attending a modernised
493 mental health day service. Mental Health Review Journal.

- 494 23 Huxley, P., Evans, S., Madge, S., Webber, M., Burchardt, T., McDaid, D., & Knapp,
495 M. (2012). Development of a social inclusion index to capture subjective and
496 objective life domains (phase II): psychometric development study. *Health*
497 *Technology Assessment*, 16(1), 1-248.
- 498 24 Freedman, V. A., Agree, E. M., Cornman, J. C., Spillman, B. C., & Kasper, J. D.
499 (2014). Reliability and validity of self-care and mobility accommodations
500 measures in the National Health and Aging Trends Study. *The Gerontologist*,
501 54(6), 944–951. <https://doi.org/10.1093/geront/gnt104>
- 502 25 Jeste, D. V., Palmer, B. W., Appelbaum, P. S., Golshan, S., Glorioso, D., Dunn, L.
503 B., & Kraemer, H. C. (2007). A new brief instrument for assessing decisional
504 capacity for clinical research. *Archives of general psychiatry*, 64(8), 966-974.
- 505 26 Revelle, W. (2020) psych: Procedures for Personality and Psychological Research,
506 Northwestern University, Evanston, Illinois, USA, [https://CRAN.R-](https://CRAN.R-project.org/package=psych)
507 [project.org/package=psych](https://CRAN.R-project.org/package=psych) Version = 2.1.3.
- 508 27 Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: The Guilford
509 Press.
- 510 28 Kracht, J. D. & Waller, N. G. (2022). Assessing Dimensionality in Non-Positive
511 Definite Tetrachoric Correlation Matrices: Does Matrix Smoothing Help?
512 *Multivariate Behavioral Research*, 57(2-3), 385-407.
- 513 29 Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal*
514 *of Statistical Software*, 48(2), 1--36. <http://www.jstatsoft.org/v48/i02/>
- 515 30 National Quality Forum (2021). *Measure evaluation criteria and guidance for*
516 *evaluating measures for endorsement*.

517 <https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=884>

518 39

519 31 Hays, W. L. (1981). *Statistics* (3rd ed.). New York, NY: Holt Rinehart Winston.

520 32 Hollander, M., Wolfe, D. A., & Chicken, E. (2014). *Nonparametric statistical methods*

521 (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

522 33 United States Census Bureau (2021, August 12). *Racial and ethnic diversity in the*

523 *United States: 2010 Census and 2020 Census.*

524 [https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-](https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html)

525 [diversity-in-the-united-states-2010-and-2020-census.html](https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html)

526

527

528

529

530

531

532

533

534

Table 1

Definitions and NQF Domains/Subdomains of Measures Selected for Development by the RTC/OM

Measure Title	The level to which individuals who use HCBS...	NQF Domain/ <i>Subdomain</i>
Meaningful Activities	engage in desired activities (e.g., education, volunteering, recreation, leisure, etc.).	Community Inclusion/ <i>Meaningful Activities</i>
Social Connectedness	develop and maintain relationships with others.	Community Inclusion/ <i>Social Connectedness</i>
Choice and Control <ul style="list-style-type: none"> ● Services and Supports ● Personal Choice ● Self-Direction 	exercise the degree of control they desire over their daily lives and the supports and services they receive.	Choice and Control/ <i>Services and Supports, Personal Choice, Self-Direction</i>
Employment <ul style="list-style-type: none"> ● Employment ● Unemployment ● Retirement 	have access to employment opportunities in line with their needs, choices, and goals, and the support necessary to maintain the desired level of employment.	Community Inclusion/ <i>Meaningful Activities/ Employment*</i>
Transportation	have timely access to appropriate transportation that supports their needs, choices and goals as related to community inclusion*.	Community Inclusion/ <i>Transportation*</i>
Freedom from Abuse and Neglect	are free from abuse and neglect and the HCBS system implements appropriate prevention and intervention strategies.	Human & Legal Rights/ <i>Freedom from Abuse and Neglect</i>

Note. Definitions taken from the NQF⁸. *Employment and Transportation were created as distinct sub-domains by the RTC-OM based on stakeholder feedback that strongly suggested they be considered as distinct important aspects of HCBS to be measured.

Table 2*Demographic Characteristics of Study Participants*

Age	Minimum	22
	Maximum	101
	M	48.3
	SD	17.8
Gender	Male	49
	Female	56
	Other/Prefer not to answer	2
Location	Rural	31
	Non-Rural	76
Disability	PD	25
	I/DD	26
	MH	26
	TBI	13
	ARD	17
Living Situation	With Family	19
	Own Home	49
	Residence with staff support	32
	Other	5
	(No Response)	2
Race	White	72
	Black/African American	18
	American Indian/Native American	3
	Hispanic/Latino	1
	Asian	1
	Native Hawaiian/Pacific Islander	1
	Other race not listed	3
	(More than one race selected)	6
Don't know/Refuse	2	
Support Source	Paid Staff	51
	Paid Family	9
	Unpaid Family	26
	Unpaid Friends	7
	Other	12
	(No Response)	2

Table 3*Reliability and Agreement Estimates for RTC/OM Measures and Subscales*

Measure	Reliability		Agreement
	Internal Consistency [95% CI*] (Sample size)	Test-retest [95% CI*] (Sample size)	Inter-observer Agreement (Sample size)
Abuse and Neglect	.62 [.51, .73] (N=100) Global [†] : .24 [-.01, .48]	.81 [.62, .91] (N=26)	.98 (N=24)
Choice and Control	.85 [.80, .89] (N=100) Global [†] : .58 [.45, .72] S&S [‡] : .77 [.70, .84] PC&G [§] : .75 [.68, .82]	.76 [.54, .88] (N=29) S&S [‡] : .74 [.51, .87] PC&G [§] : .72 [.48, .86]	.94 (N=24) S&S [‡] : .93 PC&G [§] : .98
Employment	JE : .81 [.73, .89] (N=47) B [¶] : .70 [.52, .87] (N=34)	JE : .99 [.94, 1.00] (N=10) B [¶] : .95 [.84, .99] (N=12)	JE : .89 (N=15) B [¶] : .89 (N=10)
Meaningful Activity	.94 [.92, .96] (N=101) Global [†] : .75 [.67, .84]	.79 [.59, .90] (N=29)	.92 (N=29)
Social Connectedness	.88 [.85, .92] (N=105) Global [†] : .46 [.27, .64]	.91 [.81, .96] (N=29)	.94 (N=27)
Transportation	.86 [.82, .90] (N=102) Global [†] : .75 [.67, .83]	.76 [.54, .88] (N=28)	.92 (N=28)

Note: *Confidence interval, [†]Tier 1 items only, [‡]Services & Supports, [§]Personal Choices & Goals, ^{||}Job Experiences, [¶]Barriers