

## **Supplemental Material S2.** Structure and characteristics of the adapted Communicative Participation Item Bank (CPIB; Baylor et al., 2013) scale.

Since we applied an ad-hoc modified version of the Communicative Participation Item Bank (CPIB; Baylor et al., 2013) General Short Form and people with amyotrophic lateral sclerosis (PALS) who use augmentative and alternative communication (AAC) represent a novel population for this scale (see the questionnaire and data collection section of the Methods for details), we found it necessary to perform at least a preliminary assessment of the structural coherence, cross-population generalizability, and item response characteristics of this version of the scale as we applied it. This supplemental section describes those checks.

For the purposes of this study we consider samples from three populations of interest: PALS who use only unaided communication (Group 1), and PALS who do use aided communication projected into the two scenarios of AAC availability (Group 2 with all of their communication methods) and AAC unavailability (Group 2 with unaided communication only). Note that we did not have independent samples from the latter two populations but rather one sample of the same Group 2 individuals who were asked to imagine scenarios with and without aids and respond to the CPIB items accordingly; this design feature was both a strength (we expected a low degree of confounding in the comparison between the all-methods and unaided-only conditions) and a weakness (the metric invariance of the scale could not be meaningfully tested across the latter two groups using this sample). We opted to assess scale properties across the Group 1 vs. Group 2 boundary separately in the all-methods and unaided-only conditions and consider both sets of results, recognizing that this assessment can only be viewed as tentative and that testing across independent samples from all three populations is still required for verification of the claims made below. (The two analyses can be viewed as evaluating scale properties under "easiest" and "hardest" scenarios, where in Group 1 these scenarios coincide.) Moreover, even the Group 1 and Group 2 samples at hand were recruited by convenience and each likely represented a somewhat more privileged subgroup of PALS than typical, and a subgroup with perhaps atypical facility with or access to AAC. Our modified CPIB form also included altered instructions, item phrasing, and question formatting compared to the validated General Short Form (see questionnaire and data collection section). For these reasons and others, we reiterate that the evidence concerning the scale properties of the CPIB presented here should be viewed as entirely preliminary and inadequately supportive of a recommendation for general use at this time.

The first check we undertook was investigation of the unidimensionality of the scale in the Group 1 and Group 2 populations. Cronbach's alpha was uniformly high ( $> 0.95$ , with average interitem correlations  $> 0.7$ ) in both groups in their "natural" state (without aids), and sensitivity analyses of factorizations using both iterated principal factor and maximum likelihood methods in each population separately provided strong evidence of unidimensionality, with the ratio between the first and second eigenvalue exceeding 30:1 in each case and all eigenvalues after the first being  $< 0.5$  in magnitude. The first factor explained  $> 85\%$  of the response variance. Alternative factorizations extracting  $> 2$  factors invariably lead to Heywood cases, and the (undesirable) increase in Bayesian Information

Criterion (BIC) going from 1-factor to 2-factor extractions is 7 for Group 1 and 5 for Group 2 in the unaided-only condition, suggesting that the single-factor solution is perhaps 10-40 times more likely (in a Bayesian posterior probability sense) to be the correct factorization given the response patterns in this sample. These are all strong pieces of evidence in favor of the hypothesis of unidimensionality of this version of the CPIB scale.

Interestingly, the scale when applied to Group 2 in the all-methods condition showed somewhat weaker coherence, with Cronbach's alpha 0.9 (vs.  $> 0.95$ ), average interitem correlations only 0.5 (vs.  $> 0.7$ ), and first:second eigenvalue ratio just 11:1 (vs.  $> 30:1$ ), suggesting that the addition of aids to a communication context may differ substantially in meaning from situation to situation for PALS in Group 2; this vagueness of concept may tend to decrease the coherence of item scores incorporating the benefit of aids without the additional support of qualification of which aids are being used in which communication context. Moreover, nonnegligible residual positive correlations (approaching 0.5) amongst items remained in some cases (e.g., talking with people you know was correlated with having a conversation that needs to be quick, talking with people you do not know was correlated with communicating in a small group of people, and having a conversation when upset was correlated with communicating during an emergency), suggesting that certain subgroupings of the items behaved in a correlated fashion even after adjusting responses for the scale factor, and thus carried information external to the meaning of the factor that may confound response patterns in situations where sample selection bias is operative. Another way to express this idea is that some of the items are too similar to one another to be considered fully additive bits of information. The sum of eigenvalues from factor decomposition was 7.5 instead of 10 (the number of items), implying that roughly 2.5 item units of information were lost due to overlapping nuances in the meanings of the items and replaced with unmodeled situational measurement error. But even in light of these nuances, the unidimensionality of the scale in each population was strongly supported by these initial investigations.

Following this, we investigated the consistency of item loadings on factors (i.e., metric invariance) using confirmatory factor analysis methods in a linear structural equation modeling framework. Unconstrained-coefficient models were compared to those where item loadings on the factors were constrained to be equal across Group 1 and Group 2, and the relative change in model fit was assessed by likelihood-ratio test. With the full ten-item version of the scale, fit of the constrained model was extremely poor relative to the unconstrained model in both the all-methods and unaided-only conditions, but the source of poor fit could be traced to the same two items in each case: having a short phone conversation with someone who knows you well and communicating over video calls. Not all respondents participated in these situations (see data cleaning and analysis section), calling into question the validity of extrapolating their degree of communicative participation in wholly imagined situations of this type, and certain elements in the wording (e.g., "short" and "knows you well") also seemed to risk coloring the interpretation of the situation for respondents in idiosyncratic ways. Thus, these items appeared to invite variance in responses for reasons that may have nothing to do with the trait being assessed. Psychometric properties also clearly differed across groups for these questions: Group 1 attached significantly less weight to those items than Group 2 did under either

condition, and Group 2 attached less weight to them in the all-methods condition than in the unaided-only condition. Uniqueness variance also varied strongly for these items across groups (larger for Group 2), and across the all-methods and unaided-only conditions for Group 2, pointing to situational (What aids do I use and for what purpose?) or subject-level heterogeneity (How do I feel about the aids that I use, and fare with them?) in how the items are interpreted that appeared more prominent in respondents with greater disability. For these reasons we concluded that the inclusion of the phone and video items harmed the functioning of the scale as a measurement of difficulty in communication participation, and decided to remove them when calculating total CPIB scores.

On the remaining set of eight items, metric invariance across Group 1 and Group 2 was well-supported in the all-methods condition ( $p = .33$  when comparing the constrained vs. unconstrained models) but was questionable for the unaided-only condition ( $p = .01$ ). The primary cause of the lack of fit under metric invariance in this latter scenario was that certain items with "complications" included in the item prompts (e.g., not knowing the conversation partner, when the conversation needs to be quick, etc.) were associated with significantly different weights for Group 2 participants in the all-methods and unaided-only conditions. Factor loadings for Group 2 under the all-methods condition were quite similar to those of Group 1, reasonably suggesting that the same barriers had roughly the same importance when functional ability was normalized somewhat due to use of aided communication, but the loadings of items having a burden of extra contingencies were 10% to 90% larger in magnitude for Group 2 in the unaided-only condition relative to the all-methods condition. The sharpest example of this was the item about talking with people you know, which had the smallest weight for both groups under the all-methods condition (i.e., was least relevant to the trait) but nearly double the weight for Group 2 under the unaided-only condition (i.e., was now much more relevant to the "trait," which is arguably no longer quite the same trait); the agreement in loading value with Group 1 was still fairly close, however, since the loading for this item was small to begin with (i.e., relative size of loading magnitudes reversed from Group 1 > 2 to 2 > 1 but the relative amount of divergence remained similar). So while strict metric invariance may not hold across the populations under consideration in all situations, most of the item loadings were in fact found to be very close between the groups, with large divergences being uncommon (< 15% for six of eight items) and even the largest divergences falling within 35% of agreement, suggesting that the violation would not strongly affect the summative CPIB total score. For example, assuming correlations as we see them amongst the items due to their common latent score, a 35% positive distortion in the loadings of two out of eight items would be expected to distort the eight-item total upward by approximately 5%.

Although we feel that the properties of the CPIB scale should be examined in future studies in more detail for sensitivity to complications in communication situations that may arise for PALS who need AAC but are unable to use it in the relevant contexts, we nevertheless maintain that the quantitative validity of the summative total can be considered adequate for profiling and other descriptive purposes. Note also that a common-intercepts assumption (i.e., equivalent centering of the item response scale) is highly consistent with the observations for six of eight items ( $p = .55$ ), the same items tending to show closer agreement in factor loadings between the groups in both scenarios.

(The two most problematic items related to both divergence in factor loadings and salient shifts in the center of the item scale are getting a turn in a fast-moving conversation and having a conversation when upset; both situations might have complications involving degree of disability, comfort with and reliance on AAC driving the level of frustration with its inaccessibility, and personality of the individual, as well as situational factors such as the reasons for feeling upset or for desiring to make one's voice heard in a fast-moving conversation and the perceived consequences of being unable to do so.) Thus, while this short version of the CPIB scale that we have developed for the current study should as yet be considered immature, there were still quite good indications of near-equivalence of metric properties and approximate scalar invariance across the groups we have examined.

Finally, we used item-response theory (2PL-type graded response models) to examine item-factor relationships for the eight items included in the total CPIB score. By far the item best calibrated to the total score was talking with people you do not know, which was the most discriminating item and had the highest factor loading for both groups in the all-methods condition (surpassed in importance by the two problematic items in the unaided-only condition for Group 2); this item was nearly twice as sensitive to changes in the latent difficulty score as the least-discriminating item, having a conversation that needs to be quick, for both groups in the all-methods condition (and remained twice as sensitive in the unaided-only condition for Group 2). This item (talking with people you do not know) and the item about communicating in a small group of people were both calibrated in both groups such that single-standard-deviation shifts in difficulty approximately corresponded to unit increments in the item response scale, and among all the items their thresholds covered the largest span of the latent variable range. All of the eight items included in the summative total score showed reasonable range and separation of thresholds (unlike the phone and video call items, which showed very poor separation of thresholds and indeed collapsing of adjacent response categories), always spanning at least  $> 1$  standard deviation of latent score (and often  $> 2$ ), with at least 0.5 standard deviations between thresholds. However, the mean of the latent difficulty score was quite separated for the three populations under consideration (e.g.,  $> 2$  standard deviations of latent score separation between Group 1 and Group 2 in the unaided-only condition), implying that the amount of information available for discriminating latent difficulty based on the items in the CPIB scale was very differently centered in each population. An expected total score for someone whose experience of difficulty is typical for Group 1 is about 21 (out of 24). But in Group 2 under the all-methods condition, someone experiencing average difficulty (for Group 2 in this condition) would score roughly 10, versus only 2 for someone from Group 2 who experienced average difficulty in the unaided-only condition. All the information content in Group 1 was on the negative side (distinguishing the bad-off from the worse-off), and all the information content for Group 2 in the unaided-only condition was on the positive side (distinguishing the well-off from the better-off), but for Group 2 in the all-methods condition the scale tracked degrees of difficulty with roughly equal resolution on both the positive and negative sides of typical (for that group, for that setting). As such, the scale was more sensitive to worsening conditions for someone who does not yet use aided communication, more sensitive to improving conditions for someone who needs aids in situations where they cannot or do not use them, and more sensitive to fine shades of

positive or negative change in expected conditions for someone who needs aids in situations where they do in fact use them.

In summary, we found strong evidence that this version of the CPIB scale is unidimensional and that, while strict metric invariance cannot yet be demonstrated, a summative total score based on equal loadings provides a coherent and consistent summary of the underlying construct that is reasonably sensitive to information in the scale items. However, it would be premature and imprudent to use this version of the scale more broadly without further testing on PALS populations in a broader and more controlled variety of settings involving and not involving the use of AAC. For this reason, no standard (e.g., T) scoring was attempted.

## Reference

Baylор, C., Yorkston, K., Eadie, T., Kim, J., Chung, H., & Amtmann, D. (2013). The Communicative Participation Item Bank (CPIB): Item bank calibration and development of a disorder-generic short form. *Journal of Speech, Language, and Hearing Research*, 56(4), 1190–1208. [https://doi.org/10.1044/1092-4388\(2012/12-0140\)](https://doi.org/10.1044/1092-4388(2012/12-0140))

