Predicting Optimal Preference Assessment Methods for Individuals with Developmental Disabilities

Kendra M. Thomson, Diana Czarnecki, Toby L. Martin, C.T. Yu, and Garry L. Martin
University of Manitoba and the St. Amant Research Center

Abstract: The single-stimulus (SS) preference assessment procedure has been described as more appropriate than the paired stimulus (PS) procedure for “lower functioning” individuals, but this guideline’s vagueness limits its usefulness. We administered the SS and PS preference assessment procedures with food items to seven individuals with severe or profound developmental disabilities who scored at level 2 of the Assessment of Basic Learning Abilities (ABLA) and seven who scored at ABLA level 3. Thirteen of the 14 participants also received these assessments (PS and SS), with non-food items. The two procedures were about equally effective for both groups, and with both types of stimuli, although the PS procedure produced more refined preference hierarchies. Most participants showed moderate to high correlations in preference scores between the two procedures for both food and non-food items. These results suggest that, for individuals who score at either ABLA level 2 or ABLA level 3, the SS and the PS procedures are equally likely to identify preferred stimuli.

Increasing quality of life for individuals with severe or profound developmental disabilities is an important objective for practitioners and caregivers. One strategy for increasing quality of life with this population is to assess preferences and provide them with preferred items and activities. Because many of these individuals do not have the language skills necessary to verbally communicate their preferences, various methods of non-verbal, direct preference assessment have been developed and these methods have been shown to be more reliable than caregivers’ opinions.

Two types of preference assessment are used widely: a single-stimulus (SS) procedure and a paired-stimulus (PS) procedure. In an SS procedure, a variety of stimuli (frequently 6 or 8) are presented singly and in randomized order, several times each (Pace, Ivancic, Edwards, Iwata, & Page, 1985). The experimenter prompts, “Do you want this?” and records whether or not the individual approaches each stimulus. Approach may be defined as “a voluntary movement toward the stimulus, maintaining eye contact with the stimulus for at least three seconds, exhibiting a positive facial expression, or making a positive vocalization within five seconds of the stimulus presentation” (Green, Gardner, & Reid, 1997). Typically a stimulus that has been selected or approached on at least 80% of the trials is considered a high preference stimulus (Green et al., 1997; Green et al., 1988; Green, Reid, Canipe, & Gardner, 1991; Green & Reid, 1996; Ivancic & Bailey, 1996; Logan et al., 2001; Pace et al.) and one that has been selected or approached on less than 30-50% of the trials is considered a low preference stimulus (Ivancic & Bailey, 1996; Pace et al.). The PS procedure involves presenting two stimuli concurrently on each trial and verbally prompting the individual to “Pick one.” Usually six or eight stimuli are tested and each stimulus is presented with every other stimulus an equal number of times. Observers record which stimulus the participant approaches, and compute preference scores based on a
ratio of trials that items were selected to trials that those items were available. Both methods can identify high preference stimuli that are also positive reinforcers (Green et al., 1991; Piazza, Fisher, Hagopian, Bowman, & Toole, 1996). However, the PS method differentiates preference ranks more effectively (Fisher et al., 1992).

The main limitation of the PS procedure is that it requires scanning both items, a visual discrimination that not all persons with severe and profound developmental disabilities can perform reliably. The SS procedure is therefore considered a more appropriate method for “lower functioning” individuals (Hagopian, Long, & Rush, 2004). Unfortunately, such information is not very helpful for caregivers because of the range of abilities of individuals who might be referred to as low functioning. Moreover, individuals diagnosed with severe or profound mental retardation according to DSM-IV, all of whom might be referred to as “low functioning,” show considerable variability in discrimination ability (Martin, Yu, & Vause, 2004; Richards, Williams, & Follette, 2002). Is there a more precise method for aiding caregivers to predict which of the preference assessment procedures should be used with an individual client? The Assessment of Basic Learning Abilities (ABLA; Kerr, Myerson, & Flora, 1977) may provide a solution.

The ABLA directly assesses the ease or difficulty with which a participant is able to learn a simple imitation and five, 2-choice discriminations. The imitation task (level 1) requires that a client place a manipulandum in a container after the experimenter demonstrates doing so on each trial. A two-choice position discrimination (level 2) requires an individual to place a piece of white foam in a yellow can when a red box is also present, but the can and box remain in the same position from trial to trial. A two-choice visual discrimination (level 3) requires an individual to place a piece of white foam in the yellow can when the position of the can and box are alternated randomly. A two-choice visual quasi-identity match-to-sample discrimination (level 4) requires an individual to place a small yellow cylinder in the yellow can and a red cube in the red box. Positions of the can and box alternate randomly across trials and the person is presented with either the cylinder or the cube on each trial. A two-choice auditory discrimination (level 5) requires an individual to place the white foam in either the yellow can or red box as requested verbally by the experimenter (the position of the can and box remain consistent across trials). Finally, a two-choice auditory-visual combined discrimination (level 6) requires an individual to perform the same as on level 5, however the position of the can and box are alternated randomly across trials.

Testing of a level begins with a demonstration, a guided trial, and an opportunity for an independent response. Then scoring begins. Correct responses are reinforced and incorrect responses are followed by an error correction procedure (a demonstration, guided trial, and an opportunity for an independent response). Performing eight consecutive correct responses on any ABLA level constitutes a pass on that level. The failing criterion consists of eight cumulative errors made on a specific level, including errors made on independent opportunities during error correction procedures.

Performance on the ABLA tasks is hierarchical in that an individual who fails at a given level will also fail higher levels (Martin & Yu, 2000; Martin et al., 2004). The test has high inter-tester and test-retest reliability (Martin, Yu, Quinn, & Patterson, 1983; Meyerson, 1977), and has high predictive validity for people with developmental disabilities (Martin & Yu; Martin et al., 2004). The ABLA also predicts the most effective stimulus modality for the presentation of choice options during preference assessments (Conyers et al., 2002). When given a choice between high- and low-preference edible items using object, picture, and verbal stimuli, participants chose their high-preference edible items in the modalities consistent with their ABLA results for 98% of the experimental conditions (84 of 86 phases). That is, participants who scored at level 3 were able to choose their high-preference item with objects but not pictures or spoken words. Participants who scored at ABLA level 4 identified preferences with objects and pictures, but not with spoken words. Participants at ABLA level 6 effectively identified preferences with all three stimulus modalities.

The purpose of this study was to examine
the ABLA test’s ability to predict the best preference assessment procedure with food and non-food stimuli for persons with severe or profound developmental disabilities. Specifically, we examined the performance of participants at ABLA level 2 and participants at ABLA level 3 on both the SS and the PS assessment procedures with food and non-food stimuli. We also asked a group of caregivers to predict the relative effectiveness of the SS and PS assessments with participants.

Method

Participants and Settings

Fifteen adults with severe or profound developmental disabilities were recruited from the St. Amant Centre, in Winnipeg, Manitoba, Canada, a residential and community resource facility for persons with developmental disabilities. Eight of these participants had passed up to level 2 on the ABLA test, but failed higher levels (ABLA Level 2 group in Table 1). Seven of the 15 participants had passed up to level 3, but failed higher levels (ABLA Level 3 group in Table 1). Participant H did not participate in the food assessments and Participants F and N did not participate in the non-food assessments. All ABLA tests were conducted no more than three months prior to the start of preference assessment data collection.

All sessions took place in an assessment room or in the dining room of the participant’s residence. Participants sat at a table across from the experimenter and an observer. Sessions were approximately 30 minutes in duration and each participant had approximately six to eight sessions. Sessions took place at approximately the same time of day for each participant.

For each participant, two caregivers who worked daily with that individual for at least two months (M = 8 years, 10 months) and who had not previously received training on the ABLA test were recruited to provide predictions about whether or not the SS and PS preference assessment methods would each show both a high and a low preference item.

Materials

Six food items and six non-food items were identified for each participant by the experimenters in consultation with participants’

<table>
<thead>
<tr>
<th>Participants</th>
<th>Sex</th>
<th>Age</th>
<th>Diagnosis</th>
<th>Expressive Language</th>
<th>Receptive Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABLA Level 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>M</td>
<td>34</td>
<td>Severe</td>
<td>Some requests</td>
<td>Some</td>
</tr>
<tr>
<td>B</td>
<td>F</td>
<td>36</td>
<td>Severe/Profound</td>
<td>None</td>
<td>Some</td>
</tr>
<tr>
<td>C</td>
<td>M</td>
<td>32</td>
<td>Severe</td>
<td>None</td>
<td>Some</td>
</tr>
<tr>
<td>D</td>
<td>F</td>
<td>37</td>
<td>N/A</td>
<td>None</td>
<td>Poor</td>
</tr>
<tr>
<td>E</td>
<td>M</td>
<td>31</td>
<td>Down Syndrome</td>
<td>None</td>
<td>Some</td>
</tr>
<tr>
<td>F</td>
<td>M</td>
<td>30</td>
<td>Profound</td>
<td>None</td>
<td>Some</td>
</tr>
<tr>
<td>G</td>
<td>F</td>
<td>65</td>
<td>Severe</td>
<td>None</td>
<td>Delayed</td>
</tr>
<tr>
<td>H</td>
<td>F</td>
<td>26</td>
<td>Severe</td>
<td>None</td>
<td>Some</td>
</tr>
<tr>
<td>ABLA Level 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>M</td>
<td>27</td>
<td>N/A</td>
<td>None</td>
<td>Poor</td>
</tr>
<tr>
<td>J</td>
<td>M</td>
<td>27</td>
<td>Profound</td>
<td>Unrecognized</td>
<td>Simple requests</td>
</tr>
<tr>
<td>K</td>
<td>M</td>
<td>32</td>
<td>Autistic</td>
<td>None</td>
<td>Some</td>
</tr>
<tr>
<td>L</td>
<td>F</td>
<td>49</td>
<td>Severe</td>
<td>Inappropriate</td>
<td>Some</td>
</tr>
<tr>
<td>M</td>
<td>F</td>
<td>27</td>
<td>Severe</td>
<td>None</td>
<td>Some</td>
</tr>
<tr>
<td>N</td>
<td>F</td>
<td>40</td>
<td>Severe</td>
<td>None</td>
<td>Delayed</td>
</tr>
<tr>
<td>O</td>
<td>M</td>
<td>38</td>
<td>Profound</td>
<td>None</td>
<td>Some</td>
</tr>
</tbody>
</table>

1 Diagnosis obtained from personal health records.
caregivers. The non-food items were: (a) a toy bowling ball and three pins, (b) a three-piece wooden puzzle, (c) a cloth packet stuffed with shredded cellophane that made crinkling noises when squeezed, (d) a CD player playing The Lion King soundtrack, (e) two large bouncing balls, (f) a rotating “clacker” style noisemaker, (g) a “clapper” style noisemaker much like a castanet, (h) a push-toy bear, (i) a bowl of warm water and hand lotion, (j) a paint brush, easel and paint, (k) magazines, (l) a potted artificial plant and a watering can, (m) a plush teddy bear, (n) a feather boa, (o) a movie, (p) a disco light, (q) a lightning ball, and (r) a massager. The food stimuli included snack food items such as applesauce, pretzels, cheezies, small pieces of carrot, popcorn twists, etc.

Standard ABLA test materials consisted of a red box with diagonal stripes, measuring 14 cm x 14 cm x 10 cm, a yellow can 15 cm in diameter and 17 cm in height, and an irregular shaped piece of white foam (Martin & Yu, 2000).

Paired-Stimulus Preference Assessment Procedure

During the PS assessments, two different items were presented side-by-side in front of and within the reach of the participant on each trial. The experimenter first ensured that the participant looked at each stimulus before giving the verbal prompt “pick one.” Pointing to, reaching for, or touching an item (without rejecting it by pushing it away) constituted an approach or selection response. If the participant did not touch one of the choices within 5 s for food stimuli (30 s for non-food stimuli), the trial was scored as no selection. Due to physical limitations of some participants, the trial interval was lengthened, but never exceeded 30 s.

After an item was touched, the other item was removed from sight and the experimenter said, “Here you go” and gave the selected item to the participant. The experimenter also praised the participant for choosing, and provided general encouragement after each trial (e.g., “Nice picking!” or “You are doing great!”). Selected food items were consumed. With the non-food items, the experimenter encouraged the participant to engage in the activity. Some non-food items (e.g., the magazine) were given to the participant to manipulate as he or she liked. Several non-food items were accompanied by semi-structured interaction with the experimenter (e.g., the experimenter facilitated the bowling activity with the participants).

Across 15 trials, each of the six food items was paired with every other item. The sequence was repeated with the item positions reversed in each pairing. Each item was therefore available 10 times, five times on the right and five times on the left. The same pairing process was used for the six non-food items.

Single-Stimulus Preference Assessment Procedure

During the SS assessments, only one item was presented on each trial and the experimenter asked the participant “Do you want this?” The response definitions, trial interval, and the consequences for approaching items during the SS procedure were the same as those described for the PS procedure. The six items were presented in a randomized order across trials and the sequence was repeated until each stimulus had been presented for 10 trials.

Counterbalancing

Presentation of the two formats was counterbalanced across participants. Seven participants were selected randomly to receive the PS format followed by the SS format. The other eight participants received the SS format followed by the PS format. Food and non-food items were assessed separately, and the order in which the two stimulus classes were assessed was arbitrary across participants.

Caregiver Predictions

All caregivers received written descriptions of the PS and SS procedures. They were asked to indicate which procedure would be effective in discriminating the preferences of the participants, by being able to identify high and low preference items. This was done for the food items, and then for the non-food items.

Reliability Assessments

Interobserver reliability checks were done for 25 % of each of the PS and SS preference
assessments for each participant and for each stimulus class. During reliability checks, an observer independently recorded the responses of the participants on each trial. If the same selection response was recorded by the experimenter and the observer, a trial was considered an agreement. A disagreement was recorded if the observer and experimenter recorded different selection responses. Observer agreement was calculated for each session by dividing number of agreements by number of agreements plus disagreements, and multiplying by 100% (Martin & Pear, 2003). Agreement scores averaged 98.6% across all observed sessions.

To assess procedural integrity, an observer recorded the experimenter’s behaviors on each trial using a procedural checklist (e.g., the experimenter presented correct items, the items were presented in the correct positions and the experimenter used the correct prompt, “pick one”). The procedural reliability data were collected during at least 21% of all the sessions for both the PS and SS methods with food and non-food stimuli and for each client. Procedural reliability assessment was calculated by dividing number of agreements on experimenter behaviours by total number of agreements plus disagreements during that session and multiplying by 100% (Martin & Pear). An agreement was defined as the experimenter and observer both recorded that all the procedural reliability items were administered correctly and anything else was a disagreement. Procedural reliability averaged 99.4% across participants for the food and non-food stimuli.

Results

Figure 1 shows the percentage of trials each stimulus was chosen (out of 10 trials) for each participant and for each of the SS and PS assessments with food and non-food stimuli. First, we examined whether each preference assessment procedure was able to identify at least one high (selected on at least 80% of the trials) and one low (selected on no more than 50% of the trials) preference item. For the SS method, four of the participants (A, B, D, and G) who had passed ABLA Level 2 and 4 of the participants (I, J, K, L, and M) met the high/Low criterion. The proportions of ABLA Level 2 and Level 3 participants who met the high/Low criterion with the SS method were not significantly different ($p > .05$) for either food or non-food items using the one-tailed Fisher’s Exact test. In 5 of the 14 SS food assessments (participants C, E, F, K, and O) and 3 of the 13 SS non-food assessments (participants B, G, and O), all stimuli were approached on at least 80% of the trials.

For the PS method, 4 out of 7 participants (B, C, E, and G) who passed ABLA level 2 and 3 out of 7 participants (I, J, and K) who passed ABLA level 3 met the high/Low criterion for the food stimuli. Whereas, for the non-food stimuli, 5 out of 7 participants (B, C, D, E, and H) who passed ABLA level 2 and 5 out of 6 participants (I, J, K, L, O) who passed ABLA level 3 met the high/Low criterion. The proportions of ABLA Level 2 and Level 3 participants who met the high/Low criterion were not significantly different ($p > .05$) for either food or non-food items using the one-tailed Fisher’s Exact test.

Pearson product-moment correlations between the PS and SS item preference scores ($n = 6$, for six items assessed) were computed for each participant. Mean correlation coefficients were moderate and positive, averaging .61 (ranging from −.07 to .95) for the food stimuli, and .58 (ranging from −.12 to .97) for the non-food stimuli.

Twenty-eight predictions about preference assessment method effectiveness were derived from the ABLA test. Specifically, we predicted that PS would be effective only for Level 3 participants, and that SS would be effective for both Level 2 and Level 3 participants. For the food stimuli, 50% of the 28 ABLA test predictions and 39% of the 54 caregiver predictions were accurate. For non-food stimuli, 50% of the 26 ABLA test predictions and 67% of the 52 caregiver predictions were accurate. A chi-square test of association revealed that the prediction accuracy was not significantly related to prediction source (ABLA or caregivers) ($p > .05$).
Figure 1. Percent of trials each stimulus was chosen for each participant during single-stimulus (SS) and paired-stimulus (PS) assessments with food and nonfood stimuli. Participants A through H are the ABLA Level 2 group and Participants I through O are the ABLA Level 3 group. The same stimulus in the SS and PS assessments is connected by a line.
Discussion

The main finding of the present study was that persons with severe and profound developmental disabilities, who have mastered simple two-choice position discriminations, as measured by the ABLA, could express their preferences for food and non-food items through both the SS and PS procedures. Both the SS and PS methods met the high/low criterion on 57% of the assessments in this group. The SS and PS procedures were also equally effective for participants at ABLA Level 3. Percentages of assessments that met the high/low criterion were 69% and 62% for the SS and PS procedures, respectively.

Four participants in this study showed 100% approach to all stimuli in SS: three with food stimuli and one with non-food stimuli. This is consistent with previous research that reported a tendency for participants to show a high rate of approach responses to all stimuli in the SS procedures (Fisher et al., 1992).

The lack of direct tests of item reinforcing values may have limited this study. Although participants at levels 2 and 3 of the ABLA showed preferences with both procedures, it was not shown that preferred stimuli identified by the PS method were effective reinforcers for both groups. However, preference scores are known to correlate positively with reinforcing effectiveness in teaching settings for both procedures (e.g. Green et al., 1991; Piazza et al., 1996). Furthermore, positive correlations between item preference scores given by the SS and PS procedures provide confidence that both assessments were measuring what they were intended to measure (i.e. reinforcing value). Of the 22 correlations between SS and PS procedures, 17 were above .5. Nevertheless, follow-up studies would benefit from testing both preferred and non-preferred items as reinforcers. Conducting direct preference assessment enhances the self-determination of persons who are unable to state their preferences or obtain their own reinforcers. Research on predicting the best preference assessment procedure can save time for practitioners and frontline staff. Some researchers have suggested that the SS procedure may be more appropriate for lower functioning clients (Hagopian et al., 2004). Because “lower functioning” is a vague phrase, we compared the SS and PS procedures for persons who scored at ABLA Level 2 and Level 3. Results suggest that either procedure may be viable options for both Level 2 and Level 3 individuals.

References


Martin, G. L., & Pear, J. (2003). *Behavior modification:
What is it and how to do it. Upper Saddle River, NJ: Prentice Hall.


Received: 17 September 2005
Initial Acceptance: 5 November 2005
Final Acceptance: 25 February 2006