

## Evaluation of a Tester Evaluation Form for the Assessment of Basic Learning Abilities - Revised

### Abstract

*The Assessment of Basic Learning Abilities - Revised (ABLA-R) is used to assess the learning ability of individuals with developmental disabilities. To ensure that testers administer the ABLA-R with accuracy, Martin, Martin, Yu, Thomson and DeWiele (2011) created an ABLA-R tester evaluation form (ABLA-R TEF). We evaluated the reliability and validity of the ABLA-R TEF. In Phase 1, three ABLA-R experts rated each item on the ABLA-R TEF with a mean of 6.7 on a 7-point scale, demonstrating high face validity. In Phase 2, two trained observers obtained high mean interobserver agreement of 93.9% using the ABLA-R TEF for live scoring of testers' performances while they administered the ABLA-R. In Phase 3, the ABLA-R TEF scores of testers' performances, before and after receiving training on applying the ABLA-R, were significantly different. In Phase 4, the ABLA-R TEF scores of testers were significantly correlated with subjective ratings of the testers by ABLA-R experts, demonstrating high concurrent validity. Our research indicates that the ABLA-R TEF is a valid and reliable tool for assessing the quality of an ABLA-R assessment.*

Kerr, Meyerson and Flora (1977) observed that some persons with developmental disabilities (DD) are able to learn certain tasks with ease, yet have great difficulty learning what seemed like similar tasks. They further observed that the ability to perform one or more of several standard discriminations was typically required to perform a variety of training tasks. In response to these observations they developed the Assessment of Basic Learning Abilities (ABLA) to assess whether or not individuals with DD are able to learn six discriminations, referred to as levels, including: a simple imitation, a two-choice position discrimination, a two-choice visual discrimination, a two-choice visual match-to-sample discrimination, a two-choice auditory discrimination, and a two-choice auditory-visual combined discrimination.

When assessing a tantee on the ABLA, a tester attempts to teach each of the levels, one at a time, using standardized prompting and reinforcement procedures until a pass or fail criterion on each level is met. Research has indicated that: (a) the six levels of the ABLA are ordered in difficulty from one through six with persons with DD (Kerr et al., 1977), typically developing children (Casey & Kerr, 1977), and children with autism (Ward & Yu, 2000); (b) typically developing children tend to pass all six levels by age three (Casey & Kerr, 1977); (c) the ABLA has high test-retest and inter-tester reliability (Martin, Yu, Quinn, & Patterson, 1983);

#### Authors

Nardeen Awadalla,  
Ashley L. Boris,  
Jade K. Wightman,  
Morena Miljkovic,  
Lauren Kaminski,  
Toby L. Martin,  
Garry L. Martin,  
C.T. Yu

University of Manitoba and  
St. Amant  
Winnipeg, MB

#### Correspondence

Garry.Martin@  
ad.umanitoba.ca

#### Keywords

Assessment of Basic  
Learning Abilities,  
tester evaluation form,  
developmental disabilities,  
intellectual disabilities,  
autism spectrum disorders

and (d) the ABLA can be used to match the learning ability of children and adults with DD and children with autism spectrum disorders (ASD) to the difficulty of training tasks (Martin, Thorsteinsson, Yu, Martin, & Vause, 2008; Schwartzman et al., 2009; Vause, Yu, & Martin, 2007).

Research has also indicated that the pass/fail performance of testees on ABLA Level 5 is very similar to the pass/fail performance of those testees on ABLA Level 6, and that Level 5 provides relatively little additional information regarding a testee's abilities (Martin & Yu, 2000). Additional research has indicated that a useful replacement for the original ABLA Level 5 is a two-choice visual-visual nonidentity matching (VVNM) discrimination task (Sakko, Martin, Vause, Martin, & Yu, 2004). The ABLA was recently modified to include the replacement Level 5 task, and is now called the Assessment of Basic Learning Abilities-Revised, or ABLA-R (DeWiele, Martin, Martin, Yu, & Thomson, 2011).

It is very important that direct-care staff who conduct the ABLA-R with persons with DD do so correctly as the results of the assessment may impact the type of tasks taught in a client's training program. Therefore, Martin, Martin, Yu, Thomson, and DeWiele (2011) created a tester evaluation form titled, The Assessment of Basic Learning Abilities-Revised Tester Evaluation Form (ABLA-R TEF), to assess the accuracy with which a tester administers the ABLA-R to a testee. The purpose of this research was to evaluate the validity and reliability of the ABLA-R TEF.

### **The Assessment of Basic Learning Abilities - Revised**

The six levels of the ABLA-R are presented in Figure 1. An ABLA-R assessment with a testee is usually conducted in approximately 30 minutes. Prior to assessing a level, the testee is provided with a demonstration, a guided trial and an opportunity for an independent response for that particular level. Testing of a level begins after the testee has successfully performed the independent response. Praise and an edible (e.g., a small food item preferred by the testee, such as a raisin) are given following every correct independent response. This is

done to ensure the testee's attention and effort during the assessment. Errors are followed by a three-step error correction procedure which includes a demonstration, a guided trial and an opportunity for an independent response.

Scored trials begin after the initial demonstration, guided trial and correct independent response occur. The passing criterion is eight correct responses in a row not including correct independent responses during an error correction. The failing criterion is eight total errors, including errors on an independent response as part of an error correction. Trials are repeated until the testee meets the passing criterion or the failing criterion, whichever comes first. A testee typically reaches the pass or fail criterion for any given level within 30 trials. The testee's ABLA-R level is the highest level at which they were able to achieve eight correct and consecutive independent responses.

### ***A Self-Instructional Manual for the ABLA-R***

DeWiele and colleagues (2011) prepared a self-instructional manual for the ABLA-R which is briefly described here because, as will be described later, an evaluation of the manual provided a context for our research on a tester evaluation form for the ABLA-R. The self-instructional manual includes an introduction to the ABLA-R and its levels, information on the ABLA-R materials, guidelines on how to administer the levels, study questions, and self-practice activities. There are two parts to the manual. Part 1 includes an introduction to the ABLA-R levels and testing materials, and guidelines for testing each ABLA level. In Part 2, the reader is taught how to classify different everyday training tasks according to the ABLA-R level required to perform the tasks.

### ***The ABLA-R Tester Evaluation Form***

When the ABLA-R is used to evaluate the learning ability of individuals with DD, it is very important that a tester conducts the ABLA-R assessment accurately. As stated previously, Martin and colleagues (2011) developed a tester evaluation form for the ABLA-R, referred to as the ABLA-R TEF, in order to evaluate the level of accuracy that a tester exhibits when administering the ABLA-R. The ABLA-R TEF provides

LEVEL	TEST TASK	EVERYDAY EXAMPLES
<b>Level 1 Imitation<sup>a</sup></b>	When given a piece of foam, can the student imitate the teacher placing the foam into a container?	<ul style="list-style-type: none"> <li>• Children playing Follow-the-Leader.</li> <li>• Rolling a ball from one person to the other.</li> </ul>
<b>Level 2 Position Discrimination</b>	When presented with a yellow can and a red box in a stable position, can a student consistently place a piece of foam into the container on the left?	<ul style="list-style-type: none"> <li>• Turning on the cold (vs. the hot) water tap.</li> <li>• Placing a fork on the left side of a plate when setting a table.</li> </ul>
<b>Level 3 Visual Discrimination</b>	When presented with a yellow can and a red box in randomly alternated left-right positions, can a student consistently place a piece of foam into the can?	<ul style="list-style-type: none"> <li>• Locating own printed name on the blackboard.</li> <li>• Finding a particular shirt in a closet when the location changes each time it is replaced.</li> </ul>
<b>Level 4 Visual Identity Match-to-Sample Discrimination</b>	When presented with a yellow can and a red box in randomly alternated left-right positions, can a student consistently place a small yellow cylinder into the can, and a small red cube into the box?	<ul style="list-style-type: none"> <li>• Sorting socks into pairs.</li> <li>• Restocking a partially emptied salad bar.</li> <li>• Filling containers that are partly full.</li> </ul>
<b>Level 5 Visual Non-Identity Match-to-Sample Discrimination</b>	When presented with a yellow can and a red box in randomly alternated left-right positions, can a student consistently place a purple-coloured piece of wood shaped like the word <i>Can</i> into the can, and a piece of silver-coloured wood shaped like the word <i>BOX</i> into the box?	<ul style="list-style-type: none"> <li>• Placing a cup with a saucer.</li> <li>• Putting a pencil with a piece of paper.</li> <li>• Matching the printed word CAT to a picture of a cat.</li> </ul>
<b>Level 6 Auditory-Visual Combined Discrimination</b>	When presented with a yellow can and a red box in randomly alternated left-right positions, can a student consistently place a piece of foam into the correct container when the teacher requests either “red box” or “yellow can”?	<ul style="list-style-type: none"> <li>• Responding appropriately to the spoken words “Stop” and “Go.”</li> <li>• Responding to requests such as “Stand up” vs. “sit down.”</li> </ul>

<sup>a</sup> For Level 1, researchers (e.g., Stewart, McElwee, & Ming, 2010; Martin & Yu, 2010) have debated whether “Imitation” is the best label for this level. Kerr et al. (1977) considered the names of the ABLA levels to be useful when communicating with laypeople about the ABLA.

Figure 1. The six levels of the ABLA-R and examples of everyday tasks that require mastery of the corresponding test task (reprinted with permission from DeWiele et al., 2011)

an opportunity for an observer to evaluate a tester on the proper set-up of materials for each test trial, the initial demonstration and guided trial and opportunity for an independent response, the conduct of test trials, the recording of the testee’s responses, and whether one of the pass/fail criteria was met.

The purpose of the present research was to evaluate the reliability and validity of the ABLA-R TEF. Following the model provided by Jeanson and colleagues (2010) for evaluating a behavioral checklist, we assessed the ABLA-R TEF for: (a) its face validity; (b) its inter-observer reliability for live scoring using trained observers; (c) the ability of the

ABLA-R TEF to distinguish between inaccurate and accurate administrations of the ABLA-R; and (d) its concurrent validity with ratings of testers' performance by ABLA-R experts.

## Method

### Participants and Setting

This research was conducted in conjunction with another study in which we evaluated the effectiveness of the ABLA-R self-instructional manual (Boris, 2013). Eight undergraduate university students, with no prior training in the ABLA, were recruited from a behavior modification course at the University of Manitoba. A letter was distributed by the instructor inviting the students to participate in the study indicating that participation would take the place of assigned class practica, and that participation was entirely voluntary. Students who volunteered first were chosen as participants, and informed consent was obtained prior to beginning the study.

One adult diagnosed with DD who received services from St. Amant, a residential centre for individuals with DD located in Winnipeg, Manitoba, was recruited to serve as a testee. The demographic details of the participant with DD are not described in the current study in order to preserve confidentiality. A letter was sent to the parents/legal guardian of the individual, requesting consent to participate, and outlining the nature and purpose of the study. The recruitment procedures were approved by the Psychology/Sociology Research Ethics Board at the University of Manitoba, and the ethics committee of St. Amant. All sessions were conducted in a private testing room at St. Amant.

The ABLA-R TEF was used to assess the performance of student testers as they attempted to administer the ABLA-R in the Boris (2013) study. This allowed us to evaluate the ABLA-R TEF for its inter-observer reliability for live scoring, its ability to distinguish between accurate and inaccurate administrations of the ABLA-R, and its concurrent validity, as described below. For Phases 1 and 3, three ABLA-R experts with a minimum of one year experience of conducting and/or supervising administration of the ABLA-R were recruited from St. Amant staff. Their expertise was used

to assess the face validity and concurrent validity of the ABLA-R TEF.

### Materials

The ABLA-R TEF was used to score participants' accuracy while administering the ABLA-R to a confederate role-playing a person with DD and to the participant with DD (Boris, 2013). The ABLA-R TEF consists of three sections. The first section consists of six summary sheets, one for each level of the ABLA-R. The summaries outline the steps required to test each level (see summary sheet for Level 2 in Figure 2). The second section consists of six Tester Evaluation Forms; one form for evaluating a tester while administering each level of the ABLA-R. The items on the forms assess the accuracy with which a testee completes each step required to properly administer the ABLA-R (see the Tester Evaluation Form for Level 2 in Figure 3).

### Expert's Rating Forms

To assess the face validity of the ABLA-R TEF, a rating form was developed for the experts (see the rating form for Level 2 in Figure 4). Each expert was asked to evaluate the importance of items on each Tester Evaluation Form of the ABLA-R TEF on a 7 point scale, where 1 = "not important," and 7 = "very important/essential." To assess the concurrent validity of the ABLA-R TEF, a 9-point scale, similar to the scale used by Jeanson and colleagues (2010), was used to obtain each expert's rating of each student administering the ABLA-R. The scale ranged from 1 to 3 = "poor quality ABLA Assessment," 4 to 6 = "moderate quality ABLA Assessment," and 7 to 9 = "good to very good quality ABLA Assessment."

### Procedure

#### Phase 1: Face Validity

Three ABLA-R experts were given each of the six ABLA-R TEF Tester Evaluation Forms (see the form for Level 2 in Figure 3) and a six-page questionnaire (see the questionnaire for Level 2 in Figure 4). Each page of the questionnaire asked the experts to rate each of the items for one of the Tester Evaluation Forms using a seven-point scale where 1 = "not important," 4 = "somewhat important," and 7 = "very

### **Summary of Level 2 - Position Discrimination**

#### **Initial Prompting Sequence - Don't Record Responses**

1. Place the can and box in front of the student.
2. Demonstrate. Say, "When I say, 'Where does it go?' it goes in here," while demonstrating placing the foam into the can.
3. Guided trial. Say, "Let's try together." Take the student's hand while it grasps the foam, say, "Where does it go?" and help the student to place the foam into the can. Give praise.
4. Opportunity for independent response. Say, "Now you try. Where does it go?" Give the foam to the student. If the student places the foam into the can, give praise and an edible. If the student makes an error, repeat the prompting sequence. Do not mark the data sheet.

#### **If the Student Responds Correctly on Step 4, you are Ready to Begin Scoring**

5. Say, "Where does it go?" and give the foam to the student.
6. If the student places the foam into the can:
  - Give praise and an edible.
  - Place a ✓ in the test-trials rectangle for that trial.
  - Repeat Steps 5 & 6 until the student gets 8 correct in a row.
  - Reinforce every correct response with praise and an edible.
7. If the student places the foam into the box:
  - Say, "No. That's not where it goes."
  - Shade the test-trials rectangle for that trial.
  - Do the three steps of the error correction procedure.
  - On the opportunity for an independent response, record either a ✓ or shade the error- corrections rectangle for that trial.
  - Continue error correction until a correct response occurs on an opportunity for an independent response.
  - Return to Step 5.
8. Continue until:
  - A pass occurs (8 correct test trials in a row).
  - A fail occurs (8 total errors).

*Figure 2. Summary of steps required to test Level 2 of the ABLA-R that Participants received in Baseline*

<b>ABLA-R TESTER EVALUATION FORM LEVEL 2</b>	
Scoring: For each item on each trial performed correctly, place a ✓. For items performed incorrectly, place an ✗. For items that do not apply, leave blank	
<b>Initial Prompting Sequence</b>	
1. Proper set-up	
2. Initial demonstration:	
a. "When I say, 'Where does it go?' it goes in here".	
b. Demo	
3. Initial guided trial:	
a. "Let's try together, where does it go?"	
b. Guidance	
c. Praise	
4. Initial Opp. for Ind. Res.:	
a. "Now you try. Where does it go?"	
b. Praise and edible for correct	
c. "No. That's not where it goes" for error	
d. If error: i. Demo	
ii. Guided Trial	
iii. Opp for Ind. Res.	
<b>Test Trials</b>	<b>Trial Number</b>
5. Test trials: a. "Where does it go?"	
b. Praise and reinforcer for correct	
c. "No, that's not where it goes" for error	
d. If error: i. Demo	
ii. Guided Trial	
iii. Opp. for Ind. Res.	
6. Response Recorded Immediately/Accurately	
7. Pass or fail criterion was met.	
Scoring for Level 2: Total items scored = ____ Total items scored correctly= ____	

Figure 3. Tester evaluation form for Level 2 of the ABLA-R

important." The experts were also asked to indicate if they believed that there were any items that should be added to any of the 6 Tester Evaluation Forms of the ABLA-R TEF.

**Phase 2: Inter-Observer Agreement for Live Scoring**

As described previously, the accuracy with which Participants apply the ABLA-R was assessed in an experiment (Boris, 2013) to eval-

uate the effectiveness of the ABLA-R S-IM for teaching university students how to apply the ABLA-R. The student Participants were required to: (a) attempt to administer three levels of the ABLA-R to a confederate role-playing an individual with DD after studying a brief description of the ABLA-R (Baseline); (b) master the ABLA-R S-IM, which describes the procedure for administering the ABLA-R and contains study exercises consisting of fill-in-the-blank, true/false, and multiple-choice

**Face Validity for ABLA Tester Evaluation Form Level 2**

Scoring: Please rate each item's importance in an evaluation of ABLA-R Assessments from 1-7; 1 = "not important", 4 = "somewhat important" and 7 = "very important".

Item	Rating						
	1	2	3	4	5	6	7
1. Proper set-up							
2. Initial demonstration:							
a. "When I say, 'Where does it go?' it goes in here".							
b. Demo							
3. Initial guided trial:							
a. "Let's try together, where does it go?"							
b. Guidance							
c. Praise							
4. Initial Opp. For Ind. Res.:							
a. Instruction							
b. Praise and edible for correct							
c. "No. That's not where it goes" for error							
d. If error: I. Demo							
II. Guided Trial							
III. Opp. for Ind. Res.							
5. Test trials:							
a. "Where does it go?"							
b. Praise and reinforcer for correct							
c. "No, that's not where it goes" for error							
d. If error: I. Demo							
II. Guided Trial							
III. Opp. for Ind. Res.							
6. Response Recorded Immediately/Accurately							
7. Pass or fail criterion was met.							
Are there additional items that you believe should be added to the ABLA-R TEF? If so, please comment below:							

Figure 4. Form used by the experts to evaluate the face validity of ABLA-R TEF Level 2

questions designed to help the reader learn the material, and prompts to role-play the ABLA-R with an imaginary person with DD (Training); and (c) attempt to re-administer the same three ABLA-R levels as in Baseline to a confederate

(Post-training). Participants who achieved at least 90% accuracy in conducting the ABLA-R in their Post-training assessment with a confederate proceeded to assess the ABLA-R level of an individual with DD (Generalization), in

which the Participants began testing from Level 1 and continued until the participant with DD either failed a level, or passed all six levels. Inter-observer agreement for live scoring using the ABLA-R TEF was assessed for all three assessment phases (Baseline, Post-training, and Generalization) of the Boris (2013) study.

Prior to the Boris (2013) study, two observers (undergraduate students) were trained to use the ABLA-R TEF to score an experienced ABLA-R tester (a graduate student) administering the ABLA-R to assess a confederate (another graduate student) who simulated a person with DD. The two observers had previously studied the ABLA-R SIM, and they received feedback on their practiced scoring from the first two authors and any scoring disagreements were discussed. The simulated ABLA-R sessions were video-recorded and reviewed after feedback was delivered to the observers. The observers practiced scoring using the ABLA-R TEF until they consistently reached at least 95% interobserver agreement (computed as described below).

During the Boris (2013) study, the two trained observers independently scored the accuracy with which the ABLA-R was administered by Participants using the ABLA-R TEF. Each participant attempted to administer three levels of the ABLA-R. However, different Participants attempted to administer different levels of the ABLA-R so that, across participants, the accuracy with which all ABLA-R levels were administered was assessed.

During an inter-observer agreement assessment an agreement was defined as the two skilled observers scoring an item on the ABLA-R TEF checklist identically. A disagreement was defined as the two observers scoring an item on the ABLA-R TEF differently. An inter-observer agreement was calculated for each session by dividing the number of agreements by the number of agreements plus disagreements and then multiplying by 100% (Martin & Pear, 2011).

### ***Phase 3: Use of the ABLA-R TEF to Distinguish Between Poor Quality and Good Quality ABLA-R Performance***

During the Baseline phase, the university students performed poorly while assessing a confederate role-playing a person with DD because

the students were not yet trained on the ABLA-R. During the Post-training assessment, after the students had studied the ABLA-R S-IM, it was expected that students would have the knowledge to accurately administer the ABLA-R to the confederate. The ABLA-R TEF was used to examine students' mean Baseline score compared to their mean Post-training assessment score. These scores were compared using a paired-samples *t*-test.

### ***Phase 4: Concurrent Validity***

ABLA-R TEF scores measured by trained observers were compared to subjective ratings of the testers' performances made by the two ABLA experts. During Phase 2, video clips were created for Baseline and Post-training sessions. In Phase 4 the experts were asked to score each video. To control for bias, the experts were not informed as to whether a video clip was filmed during Baseline or Post-training. The videos were scored by the experts on a 9-point scale where 1 to 3 = "poor quality ABLA-R Assessment," 4 to 6 = "moderate quality ABLA-R Assessment" and 7 to 9 = "good to very good quality ABLA-R Assessment." Once the ABLA-R experts completed their ratings, a trained observer scored the same video clips using the ABLA-R TEF. To determine the concurrent validity of the ABLA-R TEF, the degree of agreement between the ABLA-R experts' ratings and the ABLA-R TEF scores of the video clips were examined using an intra-class correlation.

## **Results**

In Phase 1, the experts rated each item of the ABLA-R TEF from 5 to 7 on the 7-point rating scale. The means for each level ranged from 6.7 to 6.8 (see Table 1) which demonstrated high face validity. In Phase 2, high inter-observer agreement was found (see Table 2). The range of average agreement was 88.6% to 99.8% for all Participants across sessions ( $M = 93.6\%$ ). In Phase 3, a paired samples *t*-test comparing good quality and poor quality administration of the ABLA-R was significant,  $t(7) = 9.56$ ,  $p < .001$ . In Phase 4, a Pearson product-moment correlation coefficient was computed to assess the relationship between the ABLA-R experts' subjective ratings of the participants' ABLA-R performance and the scores of the testers as

Table 1. Mean Face Validity Scores by Experts for the ABLA-R TEF Levels

	Expert 1	Expert 2	Expert 3	Average	Range
Level 1	6.2	7.0	7.0	6.7	5.0-7.0
Level 2	6.2	7.0	7.0	6.7	5.0-7.0
Level 3	6.3	7.0	7.0	6.8	5.0-7.0
Level 4	6.2	7.0	7.0	6.7	5.0-7.0
Level 5	6.0	7.0	7.0	6.7	5.0-7.0
Level 6	6.4	7.0	7.0	6.8	5.0-7.0
Average	6.2	7.0	7.0	6.7	-
Range	5.0-7.0	NA	NA	-	-

NA = Not Applicable

Table 2. Interobserver Agreement Scores for Live Scoring of Participants Administering the ABLA-R

Participant	Baseline 1		Baseline 2		Post-training		Generalization	
	Mean	Range	Mean	Range	Mean	Range	Mean	Range
1	90	77-97	NA	NA	99	98-100	99	99-100
2	87	77-95	90	86-97	90	82-99	NA	NA
3	94	90-98	NA	NA	93	86-99	NA	NA
4	89	75-99	97	93-100	99	99-99	98	90-100
5	92	83-98	NA	NA	NA	NA	NA	NA
6	85	67-96	81	76-91	99	98-100	97	92-100
7	97	93-100	NA	NA	95	90-99	99	99-100
8	95	92-96	87	82-91	98	97-99	98	96-100

NA = Not Applicable. Baseline 2 IOA scores for Participants 1, 3, 5, and 7 are NA as only one Baseline phase was conducted with those Participants due to the multiple-baseline design across pairs of participants. Post-training IOA scores are NA for Participant 5 as a second observer was not available. Generalization IOA scores are NA for Participants 2, 3, and 5 as these Participants were not assessed in Generalization due to performing below the mastery criterion in the Post-training assessment.

evaluated by the ABLA-R TEF. There was a significantly high correlation between the two variables,  $r = 0.8$ ,  $n = 44$ ,  $p < .001$ .

## Discussion

In Phase 1, the high importance ratings given to each of the ABLA-R TEF components by the three ABLA-R experts indicates that the ABLA-R TEF has high face validity. In Phase 2, the interobserver agreement between two trained observers while they assessed testees who used the ABLA-R to assess a confederate

role-playing a person with DD averaged 90.5% during Baseline, 96.2% during Post-training, and 98.1% during Generalization, for an overall average of 93.6% across the three phases. It should be noted that all of these averages are at or above the minimum recommended interobserver agreement of 80% for behavior modification research (Martin & Pear, 2011). The lowest individual interobserver agreement score was 67.0% and this was during a Baseline session when the participant had not yet received training on how to conduct an ABLA-R session. This may be explained by the fact that the end

of one trial and the beginning of another trial was not always apparent to the observers.

In Phase 3, the results indicate that the ABLA-R TEF effectively distinguished between good quality assessments and poor quality assessments of university students before and after receiving ABLA-R training. This is a promising result because agencies and facilities that use the ABLA-R may wish to use the ABLA-R TEF to ensure that new testers are administering the ABLA-R accurately.

In Phase 4, ABLA-R TEF scores of the students' ABLA-R performances were highly correlated with the subjective assessments of ABLA-R experts. Whether an ABLA-R score revealed a poor performance or a good performance, the experts provided similar results to that of the ABLA-R TEF using their 9-point rating scale. It is encouraging that ABLA-R TEF yielded results analogous to the ratings that experts provided, suggesting that the ABLA-R TEF has high concurrent validity.

There are several limitations of this study. First, during the initial interobserver agreement assessments of live ABLA-R sessions before Participants had received ABLA-R self-instructional training, several of the interobserver agreement scores were below 80%. The researchers experienced some difficulty in achieving high interobserver agreement scores in Baseline due to novel, unanticipated participant behaviours for which an agreement had not been defined. For instance, it was not always obvious when the participant initiated a new training trial. A second limitation of this study is that the experts who participated in Phases 1 and 4 were all employed by the St. Amant Research Centre where ABLA-R training was likely to be uniform. Future research should replicate Phases 1 and 4 with direct-care staff from other facilities. A third limitation is that our sample was limited to university students. Future research should evaluate the reliability and validity of the ABLA-R TEF with direct-care service workers to demonstrate generalizability of the results.

It is important that ABLA-R assessments are conducted accurately since the results are used to select training tasks matched to the learning ability of persons with DD, and also to track

learning progress of individuals with DD (Yu, Martin, Vause, & Martin, 2014). The results of this study suggest that the ABLA-R TEF can be used in settings such as residential facilities for individuals with DD, and thereby ensure that assessments of the ABLA-R are conducted in a consistent and accurate way so that the results obtained can be valid and useful.

## Key Messages From This Article

**Persons with disabilities:** The use of the Tester Evaluation Form for the Assessment of Basic Learning Abilities – Revised (ABLA-R TEF) can be used to ensure that direct-care staff administer the ABLA-R accurately when assessing the learning ability of persons with developmental disabilities and children with autism.

**Professionals and Policymakers:** The ABLA-R TEF can be used by behavioral researchers and by supervisors in residential facilities for individuals with developmental disabilities to ensure that assessments of the ABLA-R with persons with developmental disabilities and children with autism are conducted in a consistent and accurate way.

## Authors' Note

An expanded version of this paper was an Honours thesis submitted by the first author in partial fulfillment of the requirements of a BSc Honours degree in the Department of Psychology at the University of Manitoba.

## Acknowledgements

We would like to thank the Participants for their contribution to this project.

## References

- Boris, A. L. (2013). *An evaluation of a self-instructional manual for teaching individuals to administer the revised ABLA test to persons with intellectual disabilities*. (Master's thesis). Retrieved from <http://hdl.handle.net/1993/22201>

- Casey, L., & Kerr, N. (1977). Auditory-visual discrimination and language prediction. *Rehabilitation Psychology, 24* (monograph issue), 137-155.
- DeWiele, L., Martin, G. L., Martin, T. L., Yu, C. T., & Thomson, K. (2011). *The Kerr-Meyerson Assessment of Basic Learning Abilities Revised: A self-instructional manual* (2<sup>nd</sup> ed.). St. Amant Research Centre, Winnipeg, MB, Canada. Retrieved from <http://stamantresearch.ca/abla/>
- Jeanson, B., Thiessen, C., Thomson, K., Vermeulen, R., Martin, G. L., & Yu C.T. (2010). Field testing of the discrete-trials teaching evaluation form. *Research in Autism Spectrum Disorders, 4*, 718-723.
- Kerr, N., Meyerson, L., & Flora, J. A. (1977). The measurement of motor, visual, and auditory discrimination skills. *Rehabilitation Psychology, 24*, 95-112.
- Martin, G., Martin, T., Yu, C.T., Thomson, K., & DeWiele, L. (2011) *Assessment of Basic Learning Abilities – Revised Tester Evaluation Form*. St. Amant Research Centre, Winnipeg, MB, Canada. Retrieved from <http://stamantresearch.ca/abla/>
- Martin, G. L., & Pear, J. J. (2011). *Behavior modification: What it is and how to do it*. New York, NY: Prentice Hall/Pearson.
- Martin, G. L., Thorsteinsson, J. R., Yu, C.T., Martin, T. L., & Vause, T. (2008). The assessment of basic learning abilities test for predicting learning of persons with intellectual disabilities: A review. *Behavior Modification, 32*, 228-247.
- Martin, G. L., & Yu, C.T. (2010). A response to Stewart, McElwee, & Ming. *The Behavior Analyst, 33*, 231-233. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2995518/>
- Martin, G. L., & Yu, C.T. (2000) Overview on research of the Assessment of Basic Learning Abilities Test. *Journal on Developmental Disabilities, 7*, 10-36.
- Martin, G. L., Yu, D., Quinn, G., & Patterson, S. (1983). Measurement and training of AVC discrimination skills: Independent confirmation and extension. *Rehabilitation Psychology, 28*, 231-237.
- Sakko, G., Martin, T. L., Vause, T., Martin, G. L., & Yu, C.T. (2004). Visual-visual nonidentity matching assessment: A worthwhile addition to the assessment of basic learning abilities test. *American Journal on Mental Retardation, 109*, 44-52.
- Schwartzman, L. G., Vause, T., Martin, G. L., Yu, C.T., Campbell, L., Danbrook, M., & Feldman, M. (2009). Predicting the learning ability of children with autism: The Assessment of Basic Learning Abilities test versus parents' predictions. *Education and Training in Developmental Disabilities, 44*, 271-270. Retrieved from <http://www.daddcec.org>
- Stewart, I., McElwee, J., & Ming, S. (2010). A critical analysis of conventional descriptions of levels employed in the Assessment of Basic Learning Abilities. *The Behavior Analyst, 33*, 127-131. Retrieved from [http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2867501/#\\_\\_ffn\\_sectitle](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2867501/#__ffn_sectitle)
- Vause, T., Yu, C.T., & Martin, G. L. (2007). The assessment of basic learning abilities test for persons with intellectual disability: A valuable clinical tool. *Journal of Applied Research in Intellectual Disabilities, 20*, 483-489.
- Ward, R., & Yu, D. C.T. (2000). Bridging the gap between visual and auditory discrimination learning in children with autism and severe developmental disabilities. *Journal on Developmental Disabilities, 7*(2), 142-155.
- Yu, C.T., Martin, T. L., Vause, T., & Martin, G. L. (2014). *Kerr-Meyerson Assessment of Basic Learning Abilities Revised: Recent findings and a conceptual analysis of ordering*. Manuscript in preparation.