Detection of Differential Item Functioning in the Ethical Reasoning Identification Test

Nicole Zapparrata1, Allison Ames2, & Brian C. Leventhal1
1Center for Assessment and Research Studies, James Madison University; 2University of Arkansas

Introduction

• Ethical Reasoning in Action at James Madison University prepares enlightened citizens who apply ethical reasoning in their personal, professional, and civic lives. • Eight Key Question approach is the foundation. • Focus on assessment of ethical reasoning skills using the Ethical Reasoning Identification Test (ERIT).

• Pre-test, post-test design • ERT: 50 multiple choice items • A major threat to the validity of ERIT scores is the presence of differential item functioning (DIF) in the instrument. • DIF occurs in assessments when there is item performance difference in groups who are equal on the ability being measured, but differ due to external factors (Huggins-Manley, 2018).

• Purpose: investigate the presence of DIF in the ERIT.

• DIF in the ERIT would pose threats to validity and would require critical evaluation of the instrument (Bandalos, 2018).

Methods

Table 1
The Sample Consisting of Two Cohorts and Four Testing Occasions

<table>
<thead>
<tr>
<th>Testing Occasion</th>
<th>Sample Size</th>
<th>Average ERIT Score</th>
<th>Subgroup %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2015</td>
<td>n=465</td>
<td>34.77</td>
<td>38% Male, 76% Caucasian</td>
</tr>
<tr>
<td>Spring 2017</td>
<td>n=271</td>
<td>33.70</td>
<td>37% Male, 70% Caucasian</td>
</tr>
<tr>
<td>Fall 2016</td>
<td>n=406</td>
<td>34.56</td>
<td>39% Male, 78% Caucasian</td>
</tr>
<tr>
<td>Spring 2018</td>
<td>n=291</td>
<td>34.18</td>
<td>39% Male, 76% Caucasian</td>
</tr>
</tbody>
</table>

Note: Fall 2015 was the first cohort retested in the Spring of 2017. Fall 2016 was the second cohort retested in the Spring of 2018.

• Independent samples t-tests were used to detect impact within the subgroups for each testing occasion.

• The Mantel Haenszel procedure was used for DIF detection and converted to the ETS classification system for categorizing DIF (Zieky, 1993), which has 3 levels.

• Level-A DIF is of low concern to test developers.

• Level-B DIF is of medium concern.

• Level-C DIF is of high concern. These level-C items require critical evaluation (Bandalos, 2018).

Results

• With the exception of Fall 2016, impact was found in the average ERIT scores for both sex and ethnicity.

• After finding impact, DIF was found in twenty-nine items across the four testing occasions and subgroup analyses, of which twenty-one were ETS level-C items.

• The items favored both the reference and focal groups within each subgroup of interest.

• Table 2 summarizes the DIF findings over all testing occasions.

Table 2
ETS Level Differential Item Findings

<table>
<thead>
<tr>
<th>DIF Level</th>
<th>Sex</th>
<th>Ethnicity</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETS Level-A DIF</td>
<td>1 item detected</td>
<td>0 items detected</td>
<td>1 item</td>
</tr>
<tr>
<td>ETS Level-B DIF</td>
<td>8 items detected</td>
<td>0 items detected</td>
<td>8 items</td>
</tr>
<tr>
<td>ETS Level-C DIF</td>
<td>13 items detected</td>
<td>8 items detected</td>
<td>21 items</td>
</tr>
</tbody>
</table>

Discussion

• To make valid inferences from the ERIT, items should be invariant with respect to test-taker characteristics, beyond the construct of interest.

• Our findings detected each level of DIF present in the ERIT, with twenty-one items categorized as ETS level-C items.

• The presence of DIF threatens the validity of inferences from the ERIT.

• We recommend that Ethical Reasoning in Action discuss the level-C items found in the ERIT and critically evaluate them using sensitivity analysis.

• A content expert may provide insight into item wording features that may contribute to the presence of DIF.

• Preliminary discussions with content experts from Ethical Reasoning in Action yielded speculative causes of DIF.

• Examples of speculative causes include politics and other global influences within items.

• Some of the level-C items may need to be removed or altered.

• If the level of DIF can be decreased at the item level, the ERIT will become a more accurate measure of ethical reasoning ability and could be used to make valid inferences regarding student growth.

References

We would like to thank Ethical Reasoning in Action for their continued support and partnership with the Center for Assessment and Research Studies at JMU.