Using the Guo & Drasgow Z-Test to detect cheating in a real selection context

D. Aguado¹, A. Vidal², F. J. Abad², J. Olea², V. Ponsoda² and B. Lucía¹

¹Istituto de Ingeniería del Conocimiento (Spain)
²Universidad Autónoma de Madrid (Spain)

Introduction

1. Nowadays Internet recruitment & selection processes are widely used by most companies (Bartram, 2000; Lievens y Harris, 2003; Tippins, 2009).
2. More than 65% of the companies using Internet recruitment & selection choose Unproctored Internet Testing (UIT) as a test administration mode (Fallaw, Solomonson & McClelland, 2009).
3. Despite the large associated UIT advantages in terms of cost and flexibility, it is still a controversial practice mainly for test security and cheating reasons (Guo & Drasgow, 2010).
4. The recommendation of the International Test Commission (2006) is to conduct an evaluation in a controlled administration (CA) mode and compare UIT scores against those obtained in the CA mode, to identify cheaters. Therefore, a two-step procedure is advised (Tippins et al., 2006).
5. Different methods have been proposed to determine whether scores or answers from a person in the CA mode significantly differ from those obtained in the UIT condition, and to infer from this inconsistency the existence of cheating:
   - Person-fit methods have been used to analyze the inconsistencies in response patterns (Karabatsos, 2003)
   - Methods directly comparing scores on both applications (Guo & Drasgow, 2010; Segall, 2001)
   - Test-retest methods analyzing the item-level information (Tendeiro and Meijer, 2012)
6. There is a great interest in establishing valid methods for detecting cheating, because the consequences are relevant for both the organization and the candidate.

Research Questions: Objectives

1) The study explores the percentage of candidates flagged as cheaters by the Guo & Drasgow’s Z-test, in a real recruitment and selection context.
2) There is a problem in the validation of methods to detect cheating in empirical studies, as no information is available about who is really cheating. Therefore, research on the validation of these procedures by means of indirect evidences is important in order to validate them. Indirect evidences about the validity of the Z test is provided as well.

Method

The Test → eCat Grammar is a computerized adaptive test for assessing the proficiency in English Grammar in the Spanish population.

The Two Step-Procedure (and participants) →

<table>
<thead>
<tr>
<th>UIT</th>
<th>eCat Grammar</th>
<th>eCat Grammar (a)</th>
<th>eCat Grammar (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eCat Grammar</td>
<td>30 items</td>
<td>10 items</td>
<td>20 items</td>
</tr>
</tbody>
</table>

3506 Participants in a real selection context

The Z statistic → \( Z = \frac{\hat{\theta} - \theta_0}{\sigma} \)

Cutoff → \( Z > 2.32 \)

Results

Table 1. Z-test results

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Cheater</td>
<td>806</td>
<td>84.49</td>
</tr>
<tr>
<td>Cheaters</td>
<td>132</td>
<td>13.84</td>
</tr>
<tr>
<td>Favored</td>
<td>16</td>
<td>1.68</td>
</tr>
</tbody>
</table>

Table 2. 8 statistics: Mean and Standard Deviation

<table>
<thead>
<tr>
<th></th>
<th>UIT</th>
<th>Verification a</th>
<th>Verification b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test statistic</td>
<td>( N )</td>
<td>Test statistic</td>
<td>Test statistic</td>
</tr>
<tr>
<td></td>
<td>126</td>
<td>1.26</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>1.30</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>125</td>
<td>1.22</td>
<td>1.10</td>
</tr>
<tr>
<td>Total</td>
<td>381</td>
<td>1.29</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>0.97</td>
<td>0.87</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>1.66</td>
<td>1.54</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>0.32</td>
<td>0.20</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Significant differences:

- The mean of the estimated theta, in the UIT condition, is significantly higher in the Cheater than in the No Cheater group.
- In the Cheater group, the mean of estimated thetas in the UIT condition is significantly higher than the obtained in the verification test.

Conclusions

- The percentage of candidates flagged as cheaters by the Z test in a specific population (Spanish) and a particular test (eCat) is 13.84%.
- Some indirect evidence about the validity of the two step procedure applied has been gathered, based on:
  - The comparison of the real and simulated distributions of estimated thetas.
  - The study of response times obtained in the Cheaters and No Cheaters groups. In mean, Cheaters need 5.81 seconds more per item than No Cheaters, once item difficulty and person trait level have been partialed out.

References


Correspondence:
David Aguado
davidaguado@iic.uam.es