Perceptions of Fair Testing: What Do Candidates Think?

Presenters: Greg Pope, Questionmark
Krista Breithaupt, AICPA
Bruno D. Zumbo, University of British Columbia
Sept. 12, 2009: 10:30am to noon

Promoting Regulatory Excellence

Introduction

• Overview of project
• Data collection and methods
• Presentation of results
• Next steps
• Discussion and questions

Overview of project

• What we did:
  – Surveyed people who took, or were in the queue to take, certification or licensing exams
  – We asked 31 questions:
    • 5 demographic
    • 26 regarding what test takers think about test taking practices in a number of areas (all 4 point Agreement scale; Likert)
• Purpose:
  – To find out what candidates think about testing practices in order to help our industry understand our test takers better
Data collection and methods

- Put out the call to colleagues in the certification and licensing industry for their membership to participate in the survey
  - Nursing, Dental, Medical, Computer Science, Pharmacy, etc.
- Received a lacklustre response from colleagues
  - No organizations (save one) got back to us about participating with their membership
  - Only a little over 60 individuals took the survey despite rigorous marketing of the survey in numerous venues

- The one saving grace: American Institute of Certified Public Accountants (AICPA; thanks to Dr. Krista Breithaupt) participated fully in the survey
- AICPA sent out a newsletter to their membership with a link to the survey and a request for responses
- We obtained a fantastic response: 3586 completed responses from the AICPA membership which we will be presenting today
- We used Questionmark Perception to author and administer the surveys via internet delivery

Demographics – Gender

- Female: 1608, 45%
- Male: 1972, 55%
Demographics – Age range

Demographics – When did you take the exam

Demographics – Did you take the exam as a recertification
### Demographics – Country

<table>
<thead>
<tr>
<th>Country</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>1</td>
<td>0.03</td>
</tr>
<tr>
<td>China</td>
<td>3</td>
<td>0.08</td>
</tr>
<tr>
<td>Canada</td>
<td>4</td>
<td>0.11</td>
</tr>
<tr>
<td>Other</td>
<td>12</td>
<td>0.34</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>14</td>
<td>0.39</td>
</tr>
<tr>
<td>United States</td>
<td>3535</td>
<td>99.05</td>
</tr>
</tbody>
</table>

### Question: “I believe that it is fair for candidates to be able to retake an exam if they do not pass.”

- **99% SA or Agree**

### Question: “I believe that ‘right minus wrong’ scoring (e.g., if a candidate gets a question wrong they get a penalty of -1 towards their total exam score) arrives at fair scores for all candidates.”

- **55% SA or Agree**
Question: “I believe that exam fees are reasonable for candidates.”

467, 13% Strongly Agree
175, 5% Agree
1302, 36% Disagree
1626, 46% Strongly Disagree

51% SA or Agree

Question: “I believe that computerized adaptive tests (where different sets of questions are administered to each candidate depending on their answers to previous questions on the exam) are fair.”

1024, 29% Strongly Agree
529, 15% Agree
1677, 47% Disagree
1024, 29% Strongly Disagree

62% SA or Agree

Question: “I believe that if I need to appeal my exam score that there is a fair and accessible method for me to do so.”

2240, 54% Strongly Agree
258, 7% Agree
746, 21% Disagree
275, 6% Strongly Disagree

71% SA or Agree
Question: “I believe that the questions on the certification and licensing exams are not biased against one gender or a specific racial demographic group.”

98% SA or Agree

21.1% Agree

38.1% Disagree

14.6% Strongly Agree

20.3% Strongly Disagree

Question: “I believe that cheating on certification and licensing exams is a significant problem.”

20% SA or Agree

33.7% Agree

35.8% Disagree

19.8% Strongly Agree

10% Strongly Disagree

Question: “I believe that it is easy to obtain leaked actual exam questions on the internet if I wanted to find them.”

11% SA or Agree

35.1% Agree

35.1% Disagree

45.1% Strongly Agree

22.3% Strongly Disagree
Question: “I believe that using automatic statistical methods to detect people cheating on tests should be done more often.”

62% SA or Agree

- Strongly Agree: 124, 4%
- Agree: 1194, 34%
- Disagree: 1880, 51%

Question: “I believe that biometric authentication (e.g., fingerprint analysis) is a good way to identify candidates.”

85% SA or Agree

- Strongly Agree: 101, 3%
- Agree: 1065, 30%
- Disagree: 1952, 55%

Question: “I believe that it is acceptable to be video and/or audio recorded during an exam.”

87% SA or Agree

- Strongly Agree: 2027, 57%
- Agree: 1083, 30%
- Disagree: 331, 9%
- Strongly Disagree: 124, 4%
**Question:** “I believe that remote exam proctoring using web-cameras and audio listening devices is as effective as in-person exam proctoring methods.”

34% SA or Agree

- Strongly Agree: 669, 19%
- Agree: 1001, 28%
- Disagree: 1670, 47%
- Strongly Disagree: 224, 6%

**Question:** “If I saw someone cheating during an exam I would definitely turn them in to the exam proctor.”

80% SA or Agree

- Strongly Agree: 662, 19%
- Agree: 1074, 30%
- Disagree: 1785, 50%

**Question:** “I believe that scores that I obtain on the exam are an accurate reflection of what I know and can do.”

71% SA or Agree

- Strongly Agree: 899, 23%
- Agree: 921, 15%
- Disagree: 1979, 50%
Question: “I prefer taking computer-based exams versus paper-and-pencil exams.”

77% SA or Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

Question: “I believe that a multiple choice question format lets me demonstrate what I know and can do.”

75% SA or Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

Question: “When I am taking a certification or licensing exam I feel a great deal of stress and anxiety.”

78% SA or Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
Question: "I trust that the certification and licensing exams that I take are well constructed and provide accurate scores to candidates."

92% SA or Agree

Question: "I believe that the use of exam scores by employers/post-secondary institutions/other organizations is fair and appropriate."

69% SA or Agree

Question: "I believe that the exam scores I obtain accurately predict how well I will do in my profession (e.g., higher exam scores = higher pay or greater promotions)."

21% SA or Agree
Question: “I believe that if I took the same certification or licensing exam again that I would obtain a similar score.”

63% SA or Agree

Question: “I believe that the availability of practice tests is essential to my performing well on the actual certification or licensure exam.”

96% SA or Agree

Question: “I believe that taking practice tests before I take the actual certification or licensing exam reduced my stress and anxiety when taking the actual exam.”

89% SA or Agree
Question: "I believe that the reports I obtain regarding my performance on the exam are clear and concise."

72% SA or Agree

Question: "I believe that as a candidate that organizations/companies that create certification and licensing exams have my best interest in mind."

75% SA or Agree

Question: "I believe that certification and licensing exams are necessary to ensure the quality of candidates in my industry."

90% SA or Agree
### Slicing and Dicing: Gender by questions (data)

<table>
<thead>
<tr>
<th>Question</th>
<th>Relationship</th>
<th>Question</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fair to retake</td>
<td>0.059*</td>
<td>Scores reflect</td>
<td>-0.057*</td>
</tr>
<tr>
<td>Right minus wrong</td>
<td>-0.025*</td>
<td>Prefer CBT</td>
<td>-0.044*</td>
</tr>
<tr>
<td>Exam fees</td>
<td>-0.184*</td>
<td>MC format</td>
<td>-0.023</td>
</tr>
<tr>
<td>CAT fair</td>
<td>-0.108**</td>
<td>Stress and anxiety</td>
<td>0.273**</td>
</tr>
<tr>
<td>Appeal exam</td>
<td>0.047*</td>
<td>Well constructed</td>
<td>-0.033</td>
</tr>
<tr>
<td>Item bias</td>
<td>0.132**</td>
<td>Use of scores</td>
<td>-0.020</td>
</tr>
<tr>
<td>Cheating problem</td>
<td>0.007**</td>
<td>Scores predict</td>
<td>-0.003</td>
</tr>
<tr>
<td>Obtain leaked items</td>
<td>0.017</td>
<td>Test retest similar</td>
<td>-0.024*</td>
</tr>
<tr>
<td>Stat cheater detect</td>
<td>0.025**</td>
<td>Practice test perform</td>
<td>0.077**</td>
</tr>
<tr>
<td>Biometrics</td>
<td>-0.053**</td>
<td>Practice test stress</td>
<td>0.020</td>
</tr>
<tr>
<td>Video/audio</td>
<td>0.024**</td>
<td>Reports clear</td>
<td>0.091*</td>
</tr>
<tr>
<td>Remote proctor</td>
<td>-0.002</td>
<td>Best interests in mind</td>
<td>0.033*</td>
</tr>
<tr>
<td>Turn in cheater</td>
<td>0.043**</td>
<td>Exams are necessary</td>
<td>-0.075**</td>
</tr>
</tbody>
</table>

Gender coded as 1=Male, 2=Female; Questions coded 4=Strongly Agree, 1=Strongly Disagree

*Statistically significant to 0.05, **Statistically significant less than or equal to 0.01

### Slicing and Dicing: Gender by questions (translation)

<table>
<thead>
<tr>
<th>Question</th>
<th>Gender agreeing more</th>
<th>Question</th>
<th>Gender agreeing more</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fair to retake</td>
<td>Females (small diff)</td>
<td>Scores reflect</td>
<td>Males (small diff)</td>
</tr>
<tr>
<td>Right minus wrong</td>
<td>Males (small diff)</td>
<td>Prefer CBT</td>
<td>Males (small diff)</td>
</tr>
<tr>
<td>Exam fees</td>
<td>Male (small diff)</td>
<td>MC format</td>
<td>No difference</td>
</tr>
<tr>
<td>CAT fair</td>
<td>Male (small diff)</td>
<td>Stress and anxiety</td>
<td>Females (med diff)</td>
</tr>
<tr>
<td>Appeal exam</td>
<td>Females (small diff)</td>
<td>Well constructed</td>
<td>No difference</td>
</tr>
<tr>
<td>Item bias</td>
<td>Male (small diff)</td>
<td>Use of scores</td>
<td>No difference</td>
</tr>
<tr>
<td>Cheating problem</td>
<td>Females (small diff)</td>
<td>Scores predict</td>
<td>No difference</td>
</tr>
<tr>
<td>Obtain leaked items</td>
<td>No difference</td>
<td>Test retest similar</td>
<td>Male (small diff)</td>
</tr>
<tr>
<td>Stat cheater detect</td>
<td>No difference</td>
<td>Practice test perform</td>
<td>Female (small diff)</td>
</tr>
<tr>
<td>Biometrics</td>
<td>Male (small diff)</td>
<td>Practice test stress</td>
<td>No difference</td>
</tr>
<tr>
<td>Video/audio</td>
<td>Male (small diff)</td>
<td>Reports clear</td>
<td>Females (small diff)</td>
</tr>
<tr>
<td>Remote proctor</td>
<td>No difference</td>
<td>Best interests in mind</td>
<td>Females (small diff)</td>
</tr>
<tr>
<td>Turn in cheater</td>
<td>Female (small diff)</td>
<td>Exams are necessary</td>
<td>Male (small diff)</td>
</tr>
</tbody>
</table>

A positive correlation means that females agreed more with the question
A negative correlation means that males agreed more with the question

---

### Gender analysis: “When I am taking a certification or licensing exam I feel a great deal of stress and anxiety.”

![Gender analysis graph](image)
Gender analysis: “I believe that the questions on the certification and licensing exams are not biased against one gender or a specific racial or demographic group.”

Slicing and dicing: Age range (data)

<table>
<thead>
<tr>
<th>Question</th>
<th>Relationship</th>
<th>Question</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fair to reliable</td>
<td>0.42*</td>
<td>Scores reflect</td>
<td>-0.035*</td>
</tr>
<tr>
<td>Right minus wrong</td>
<td>0.001</td>
<td>Prefer CBT</td>
<td>-0.036**</td>
</tr>
<tr>
<td>Exam fees</td>
<td>0.231**</td>
<td>MC format</td>
<td>-0.197*</td>
</tr>
<tr>
<td>CAT fair</td>
<td>0.062**</td>
<td>Stress and anxiety</td>
<td>0.000</td>
</tr>
<tr>
<td>Appeal exam</td>
<td>0.060**</td>
<td>Well constructed</td>
<td>0.006</td>
</tr>
<tr>
<td>Item bias</td>
<td>-0.294**</td>
<td>Use of scores</td>
<td>-0.220**</td>
</tr>
<tr>
<td>Cheating problem</td>
<td>0.027</td>
<td>Scores predict</td>
<td>0.013</td>
</tr>
<tr>
<td>Obtain leaked items</td>
<td>0.106**</td>
<td>Test retest similar</td>
<td>-0.153**</td>
</tr>
<tr>
<td>Stat cheater detect</td>
<td>0.067**</td>
<td>Practice test perform</td>
<td>0.018</td>
</tr>
<tr>
<td>Biometrics</td>
<td>-0.003</td>
<td>Practice test stress</td>
<td>-0.024**</td>
</tr>
<tr>
<td>Video/audio</td>
<td>-0.180**</td>
<td>Reports clear</td>
<td>0.010</td>
</tr>
<tr>
<td>Remote proctor</td>
<td>0.275*</td>
<td>Best interests in mind</td>
<td>-0.044**</td>
</tr>
<tr>
<td>Turn in cheater</td>
<td>0.020</td>
<td>Exams are necessary</td>
<td>0.020</td>
</tr>
</tbody>
</table>

Age range coded as: 1=18-21, 7=61 or older; Questions coded 4=Strongly Agree, 1=Strongly Disagree
*Statistically significant to 0.05, **Statistically significant less than or equal to 0.01

Slicing and dicing: Age range (translation)

<table>
<thead>
<tr>
<th>Question</th>
<th>Age agreeing more</th>
<th>Question</th>
<th>Age agreeing more</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fair to reliable</td>
<td>Older (small diff)</td>
<td>Scores reflect</td>
<td>Younger (small diff)</td>
</tr>
<tr>
<td>Right minus wrong</td>
<td>Older (small diff)</td>
<td>Pref CBT</td>
<td>Younger (small diff)</td>
</tr>
<tr>
<td>Exam fees</td>
<td>Older (small diff)</td>
<td>MC format</td>
<td>Younger (small diff)</td>
</tr>
<tr>
<td>CAT fair</td>
<td>Younger (small diff)</td>
<td>Stress and anxiety</td>
<td>No difference</td>
</tr>
<tr>
<td>Appeal exam</td>
<td>Older (small diff)</td>
<td>Well constructed</td>
<td>No difference</td>
</tr>
<tr>
<td>Item bias</td>
<td>Younger (small diff)</td>
<td>Use of scores</td>
<td>Younger (small diff)</td>
</tr>
<tr>
<td>Cheating problem</td>
<td>No difference</td>
<td>Scores predict</td>
<td>No difference</td>
</tr>
<tr>
<td>Obtain leaked items</td>
<td>Older (small diff)</td>
<td>Test retest similar</td>
<td>Younger (small diff)</td>
</tr>
<tr>
<td>Stat cheater detect</td>
<td>Older (small diff)</td>
<td>Practice test perform</td>
<td>No difference</td>
</tr>
<tr>
<td>Biometrics</td>
<td>No difference</td>
<td>Practice test stress</td>
<td>Older (small diff)</td>
</tr>
<tr>
<td>Video/audio</td>
<td>Younger (small diff)</td>
<td>Reports clear</td>
<td>No difference</td>
</tr>
<tr>
<td>Remote proctor</td>
<td>Older (small diff)</td>
<td>Best interests in mind</td>
<td>Younger (small diff)</td>
</tr>
<tr>
<td>Turn in cheater</td>
<td>No difference</td>
<td>Exams are necessary</td>
<td>No difference</td>
</tr>
</tbody>
</table>

A positive correlation means that older test takers agreed more with the question
A negative correlation means that younger test takers agreed more with the question
Age range analysis: “I believe that exam fees are reasonable for candidates.”

Age range analysis: “I prefer taking computer-based exams versus paper-and-pencil exams.”

Age range analysis: “I believe that the use of exam scores by employers/post-secondary institutions/other organizations is fair and appropriate.”
Next steps

- More slicing and dicing the data, working with AICPA on follow-up studies
- Expanding the study to more test takers in more/different areas (e.g., nursing, medical)
  – Interesting in being part of this? Let us know!
- Expanding the study to different areas. For example: “What employees at testing organizations think”

Discussion and questions

- Thank you for attending
- Discussion and Questions?

Speaker Contact Information

- Greg Pope, Analytics and Psychometrics Manager, Questionmark
  - greg.pope@questionmark.com
- Krista Breithaupt, Director of Psychometrics and Research, AICPA
  - kbreithaupt@aicpa.org
- Bruno D. Zumbo, Professor of Measurement and Statistics, University of British Columbia (UBC)
  - bruno.zumbo@ubc.ca