

A Review of *Integrity*TM

Bernard P. Veldkamp
University of Twente
The Netherlands

*Integrity*TM, an online application for testing both the statistical integrity of the test and the academic integrity of the examinees, was reviewed. Program features, and the program output are described. An overview of the statistics in *Integrity*TM is provided. The application is illustrated with a small simulation study. The application is very user friendly, fast, and provides many useful statistics. Program requirements and purchasing information are provided.

Key words: classical item and test analysis, collusion detection, detection of answer copying, online application

Introduction

At all levels of schooling, teachers face the problem of cheating. Several surveys revealed that up to eighty percent of students admit cheating at least once during their career. It turns out that it is very hard to fully prevent cheating. The temptation to cheat, especially in high-stakes testing is great, and many students seem to view cheating as one of the possible strategies to succeed on tests. However, reporting scores that have been obtained by cheating might cause serious problems for the validity of testing programs. In the NCME instructional module on quality control procedures in the scoring, equating, and reporting of test scores, Allalouf (2007) stated that educational institutions should do their utmost best to ensure that a test score represents the honest achievement of an examinee.

The most common ways of cheating are impersonation, answer copying, knowing the test items and answers in advance, communicating with others during the test, bringing prohibited materials, and receiving help from proctors. Several steps can be taken to prevent and detect these kinds of fraud (for a comprehensive review of preventive measures and a large number of examples, see Cizek, 1999). To detect/prevent answer copying, statistical procedures are available. They can be used either to detect aberrant response patterns, or to detect unexpected high similarities between answers of students. *Integrity*TM (see Table 1 for product information) is an online

application that can be used to analyze tests and that offers five statistics to detect unexpected high similarities between response vectors.

Program description and features

Integrity™ is an online application that can be used to analyze multiple-choice data. Users upload their data files (in ASCII-format) on a secure server and Integrity™ provides them with a number of reports on both the statistical integrity of the test and the academic integrity of the examinees. These detailed reports can be used either to improve the test and the test items, or to confront examinees with their aberrant behavior. In the reports, users get statistical information about the performance of the test and the examinees. Supporting information on how to interpret the reports can be found at the website.

When users upload their data, a wizard guides them through the process of selecting the kind of analyses to be performed. Classical item and test statistics are automatically generated, but users can indicate whether collusion detection analyses, analyses of subscales, or analysis of subgroup behavior should be carried out. Advanced users are even allowed to modify the settings of the collusion statistics. Once the data are uploaded, the application reports the progress of the analyses online. Users can also check the performance of the Integrity™ server on the online server performance indicator.

The output of the programs consists of a number of web-based reports that can be downloaded in PDF format. An Executive summary report gives a quick overview of the key findings. Reports at the item and test levels about collision detection, subgroup and subscale analyses, and individual examinee results provide information about the various types of analyses that have been carried out. In addition, data files that contain all statistical information produced by Integrity™ can be generated.

TABLE 1
Summary Information on Program Requirements and Availability

Program name	Integrity™
Vendor	Castle Rock Research Corp, 2340 Manulife Place, 10180-101 Street, Edmonton, Alberta, Canada T5J 3S4. Tel: (780) 448 9619, Fax (780) 426 3917, Toll Free 1-800-840-6224. E-mail: integrity@castlerockresearch.com. Web: integrity.castlerockresearch.com
Price	Different licenses are available. The price depends on the number of examinees. The smallest package, for analyzing 10,000

	examinees, costs \$ 2,000.
System requirements	Microsoft Windows 98/ME/2000/XP; Pentium III processor or AMD processor at 500 MHz or greater; 128 MB RAM; 1024 by 768 pixel display with 32 bit color; Cable internet connection (293.0 KB/s) or faster.
Browser requirements	MAC OS/X 10.3.5; PowerPC G4 (2.1) 400 MHz or greater Internet Explorer 6.0 or later; Mozilla Firefox 1.0 or later; Netscape 7.2 or later.

Item and test statistics

Integrity™ has implemented a number of classical item and test statistics. At the item level, the difficulty (p-value), the discrimination (corrected point-biserial correlation coefficient), and the KR-20 are calculated. The proportion of examinees selecting each of the alternatives is calculated for the whole group and for sub-groups of low, medium and high performing examinees. Finally, the average total test score for the examinees that selected an alternative is reported. At the test level, a number of descriptive statistics such as mean score, standard deviation, skewness, kurtosis, and minimum and maximum score are given in addition to several reliability measures. When a test consists of several subscales, the correlations between the subscales are also presented.

Statistics for collusion detection

Five different methods for detection of answer copying have been implemented in Integrity™:

Angoff's B-index	Statistic based on identifying the number of matching incorrect responses (Angoff, 1974).
ACT Pair 1 method	Statistic that not only focuses on the number of matching incorrect scores, but also takes the length of the longest string of identical answers into account (Hanson, Harris, & Brennan, 1987).
ACT Pair 2 method	Statistic that uses both the number of matching incorrect answers in the longest string of identical answers and a statistic based on the number of items, the number of identical responses, and the number of identical incorrect responses (Hanson, Harris, & Brennan, 1987).
MESA method	Statistic that computes the probability that a pair of examinees has a certain number of matching incorrect answers (Modified version of

Belezza, & Belezza, 1991)

G2 method

The probability is calculated that the copier selects the same alternative as the source, given the popularity of each alternative and the copiers' total score (Frary, Tideman, & Watts, 1977).

For each pair of examinees, Integrity™ calculates whether these statistics are significant. Three different alpha levels are being used, and the outcomes are denoted as low, medium, or highly significant. I did not find references to the actual alpha levels implemented. The users are encouraged to check for collusion as soon as a pair of examinees is flagged and statistical significance is found for at least one of the statistics. Integrity™ also provides the option of comparing the results of this examinee with all the others or just to examine a small group of examinees, in case suspicion exists that one or some of the examinees have copied.

Group statistics

One of the options in Integrity™ is to compare the performance of different subgroups of examinees, for example gender, age group, or writing center. A maximum of two group indicators can be added to the data file. For each subgroup the test statistics will be calculated, and graphs of cumulative percent per group and comparison of group means are reported. Although no formal DIF analysis is carried out, these statistics can provide some information about how the test performs for different subgroups of examinees.

Documentation

The documentation for Integrity™ consists of a number of online documents. No formal manual is available. In the Knowledge Base section of the application, several Quick Start guides are provided in PDF format. They are well written short documents that explain how to upload files, and how to interpret the different reports. In addition, background documents on how Integrity™ can help instructors in the classroom, how it can help evaluate and address issues of academic integrity, and how it can help to improve the quality of assessments can be downloaded. Finally, a list of suggested readings is provided. All these documents have been written to provide the users with enough information to be able to understand and use the application, without having to study all procedures in detail. For more detailed information, references to academic sources on the different topics are included in the documents.

Program illustration

In order to illustrate the performance of the application a small simulation study was carried out. For a small test (n = 25 items) and for a longer test (n = 60 items) the collusion statistics were calculated whereby one pair of examinees copied between 0% and 100% of their answers. The source answered fifty percent of the items correctly. The tests were administered to 225 and 215 examinees, respectively. For different percentages of copying, it is shown in Table 2 and Table 3 whether copying was detected with low, moderate, or high statistical confidence. It took Integrity™ between five and eleven seconds to complete the analyses for these datasets.

Table 2

Statistical confidence for detecting answer copying for a small test (n=25)

Statistic	Percentage copied											
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	
B-index	-	-	-	-	-	-	-	-	-	-	-	-
Pair 1	-	-	-	-	-	-	-	-	-	-	-	low
Pair 2	-	-	-	-	-	-	-	low	mod	high	high	high
MESA	-	-	-	-	-	-	-	-	low	low	low	low
G2	-	-	-	-	-	-	-	-	-	low	mod	mod

Table 3

Statistical confidence for detecting answer copying for a longer test (n=60)

Statistic	Percentage copied											
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	
B-index	-	-	-	-	-	low	mod	high	high	high	high	high
Pair 1	-	-	-	-	-	low	mod	mod	mod	high	High	High
Pair 2	-	-	-	-	low	high	high	high	high	high	High	High
MESA	-	-	-	-	-	-	-	low	mod	mod	high	high
G2	-	-	-	-	-	-	mod	high	high	high	High	High

For the small test, copying was detected with a low statistical confidence when 70% of the answers were copied. For the longer test, copying was detected when 40% of the answers were copied. These results illustrate the observation by Sotaridona and Meijer (2002) that the power of most of the statistics for detection of answer copying is rather low. Examinees have to copy many (incorrect) answers in order to be detected. In the Quick Starting guide for interpreting the results of collusion analysis, it is stated that rather conservative bounds for detection of answer copying are implemented in Integrity™. Therefore, the recommendation that an examinee pair has to be

scrutinized when one of the statistics flag a pair of examinees as potential answer copiers, should be taken seriously.

Discussion and evaluation

In this review, I mainly focused on the collision analysis features of Integrity™. The reason is that several tools for performing classical item and test analysis are available, but the feature of collision analysis distinguishes this application from the other tools. Even though it remains difficult to detect answer copying, performing collusion analysis and broadcasting among students that these analyses will be carried out will increase the security and the validity of any testing program.

My only point of criticism of Integrity™ is related to the choice of statistics. The most recent statistic that is implemented was developed in 1991. In the past fifteen years, several other statistics have been developed and thoroughly researched. The kappa-index (Holland, 1996, Sotaridona, & Meijer, 2002, van der Linden, & Sotaridona, 2004) has received a lot of attention, and several IRT-based statistics were proposed (Wollack, 1997, van der Linden, & Sotaridona, 2006). Implementing some of these more recent statistics might further increase the power of Integrity™. A further suggestion would be to add statistics to the program that check for aberrant response patterns.

Despite these remarks, Integrity™ is very user friendly, fast, and provides users with many useful statistics. Allalouf (2007) already suggested that collision analysis should be performed on a routinely basis by every testing organization. Integrity™ might just be the tool for that.

References

- Allalouf, A. (2007). Quality Control Procedures in the Scoring, Equating, and Reporting of Test Scores. *Educational Measurement: Issues and Practice*, 26, 36-43.
- Angoff, W. (1974). The development of statistical indices for detecting cheaters. *Journal of the American Statistical Association*, 69, 44-49.
- Belezza, F., & Belezza, S. (1991). Detection of cheating in multiple choice tests by using error-similarity analysis. *Teaching of Psychology*, 16, 151-155.
- Cizek, G.J. (1999). *Cheating on Tests, How to Do It, Detect It, and Prevent It*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Frary, R.B., Tideman, T.N., & Watts, T.M. (1977). Indices of cheating on multiple-choice tests. *Journal of Educational Statistics*, 2, 235-256.

- Hanson, B.A., Harris, D.J., & Brennan, R.L. (1987). *A comparison of several statistical methods for examining allegations of copying*. (ACT Research Report Series No 87-15). Iowa City, IA: American College Testing.
- Holland, P. W. (1996). *Assessing unusual agreement between the incorrect answers of two examinees using the K-index: Statistical theory and empirical support (Research Report RR 96-7)*. Princeton, NJ: Educational Testing Service.
- Sotaridona, L. S., & Meijer, R. R. (2002). Statistical properties of the K-index for detecting answer copying. *Journal of Educational Measurement*, 39, 115-132.
- van der Linden, W. J., & Sotaridona, L. S. (2004). A statistical test for detecting answer copying on multiple-choice tests. *Journal of Educational Measurement*, 41, 361-377.
- van der Linden, W. J., & Sotaridona, L. S. (2006). Detecting Answer Copying When the Regular Response Process Follows a Known Response Model. *Journal of Educational and Behavioral Statistics*, 31, 283-304
- Wollack, J. A. (1997). A nominal response model approach to detect answer copying. *Applied Psychological Measurement*, 21, 307-320.