Eurasian Journal of Educational Research 74 (2018) 207-226

Eurasian Journal of Educational Research
www.ejer.com.tr

Investigating the Performance of Omega Index According to Item Parameters and Ability Levels*

Onder SUNBUL[1] , Seha YORMAZ[2]

ARTICLE INFO

Article History:
Received: 18 Dec. 2017
Received in revised form: 01 Feb. 2018
Accepted: 09 Mar. 2018
DOI: 10.14689/ejer.2018.74.11

Keywords
Answer copy detection, cheating, test security.

ABSTRACT

Purpose: Several studies can be found in the literature that investigate the performance of ω under various conditions. However no study for the effects of item difficulty, item discrimination, and ability restrictions on the performance of ω could be found. The current study aims to investigate the performance of ω for the conditions given below. **Research Methods:** b parameter range was restricted in two levels (-2.50 – 0.00, 0.01 – 2.50); a parameter range, in two levels (0.10 – 0.80 and 0.81 – 1.50). After crossing a and b parameter ranges, four different item parameter cells were obtained. 10,000 examinee responses were generated for each item parameter cell for 20 items. After combining four data sets, an 80-itemdataset was obtained. In order to obtain the effects of source's and copier's ability levels to the performance of ω, ability range was divided into four intervals (-3.00 – -1.50, -1.50 – 0.00, 0.00 – 1.50 and 1.50 – 3.00). By crossing the ability ranges of source and copier, sixteen different combinations were obtained. Each of the sixteen ability pairs of source and copier cheating was investigated for item parameter crossing cells for power study of ω. For Type I error study, no cheating data were investigated for the same conditions and levels. **Findings:** Type I error inflations were observed for the lower copier ability levels. The results of the power study indicate that when high ability level copier copied answers of the low difficulty level and high discriminative items from high ability level source, power of ω was weakened. **Implications for Research and Practice:** The study suggests that researchers must pay attention to copiers - source ability level and copied items' difficulty levels while using ω index for detecting answer copying.

*This study is the revised and improved version of the paper presented at Conference on Test Security in Madison, Wisconsin, 6-8 September, 2017.

[1]Mersin University, TURKEY, e-mail: ondersunbul@gmail.com, ORCID ID: orcid.org/0000-0002-1775-1404
[2]Mersin University, TURKEY, e-mail: sehayormaz@gmail.com, ORCID ID: orcid.org/0000-0002-8385-3724
Corresponding Author: Onder SUNBUL, Faculty of Education, Department of Measurement and Evaluation in Education, Mersin University, e-mail: ondersunbul@gmail.com, ORCID ID: orcid.org/0000-0002-1775-1404

Introduction

Multiple-choice items are frequently used for high-stakes examinations because of their particular advantages. Like any other examinations, reliability and validity of multiple-choice item tests are vital for making decisions about examinees. However, there are several threats to reliability and validity. One threat to validity is cheating. It is known that examinees often cheat in examinations. Technological developments, accelerated during the 21th century, combined with the creativity of examinees cause new ways of cheating in examinations. A subfield of measurement of test security domain upgraded itself as a reaction to new cheating ways, to protect the examination from cheating attempts before, during or after testing (mostly relevant with test security parts of testing organization). To escape unintended results of testing, testing organizations must be very careful about test security issues. The cost of weakness in test security will be very high for testing organizations. Despite test security sensitivity, testing organizations will still come across cheating. It is possible to detect cheating after examination in three distinct parts (Cizek and Wollack, 2017). The first part relates to answer copying - similarity – aberrance; the second part is item pre-knowledge; and the last part is unusual score gain and erasure detection.

Since systematic cheating is a big threat to validity, through the decades numerous methods have been developed to detect cheating (Bird,1927, 1929; Anikeef, 1954; Saupe, 1960; Angoff, 1974; Frary, 1993; Frary, Tideman and Watts, 1977; Hanson, Harris and Brennan, 1987; Bellezza and Bellezza, 1989; Bay, 1994; Harpp, Hogan and Jennings, 1996; Holland, 1996; Wollack, 1997; Wesolowsky, 2000; van Krimpen-Stoop and Meijer, 2001; Sotaridona and Meijer, 2002, 2003; van der Linden and Sotaridona, 2004, 2006; Giardano, Subhiyah and Hess, 2005; Sotaridona, van der Linden and Meijer, 2006; van der Ark, Emons and Sijtsma, 2008; Deng, 2008; Armstrong and Shi, 2009; Maynes, 2009; Belov and Armstrong, 2010; Clark, 2010; Hui, 2010; Belov, 2011; Shu, 2011; Wollack and Maynes, 2016). It should be noticed that the assumptions of these statistics are very important for their performance.

This study is related to the ω (omega) index, which is one of the most frequently-used answer copying statistics, due to its performance. Theoretical background of ω is given below:

ω Index

ω index computes the similarity of answer vectors of a given pair by comparing an observed match with an expected match (Wollack, 1997). ω index uses Nominal Response Model of Item Response Theory for probability calculations, which is required to obtain an expected match. It is possible to use Nominal Response Model to calculate the probability of selecting an option from a multiple-choice item for a given ability level.

To obtain the observed match of a given pair of suspected copier (C) and suspected source (S):

$$h_{CS} = \sum_{i=1}^{n} I[u_{iC} = u_{iS}]$$

where

h_{CS}: Observed Match

n: Number of Items

I: In the case of match=1 else 0

u_{iC}: Response of suspected copier to the ith item

u_{iS}: Response of suspected source to the ith item

To obtain the expected match of a given pair of suspected copier (C) and suspected source (S):

$$E(h_{CS}|\theta_C, U_S, \xi) = E\left[\sum_{i=1}^{n} I(u_{iC} = u_{iS}|\theta_C, U_S, \xi)\right] = \sum_{i=1}^{n}[P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)]$$

where

$E(h_{CS}|\theta_C, U_S, \xi)$: Expected match

I: In the case of match I=1 else I= 0

u_{iC}: Response of suspected copier to the ith item

u_{iS}: Response of suspected source to the ith item

θ_C: Suspected copiers ability estimate

ξ: Item parameters (a and b for dichotomous responses, lambda and zeta for nominal responses)

Depending on the local independence assumption of Item Response Theory, h_{cs} is the sum of independent Bernoulli variables under the estimation of copier ability level, source ability level, and item parameters (Wollack, 1997). The probability of responses that are copied from source by copier and its variance are given below:

$$P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)$$

and

$$[P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)][1 - P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)]$$

When the number of items approaches infinity, h_{cs} tends to show normal distribution, and ω distribution approaches standard normal distribution, for a given copier and source (Wollack, 1997).

Formula of ω is given below:

$$\omega = \frac{h_{CS} - E(h_{CS}|\theta_C, U_S, \xi)}{\sigma_{h_{CS}} - E(h_{CS}|\theta_C, U_S, \xi)} = \frac{h_{CS} - \sum_{i=1}^{n}[P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)]}{\sqrt{\sum_{i=1}^{n} P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)[1 - P(u_{iC} = u_{iS}|\theta_C, U_S, \xi)]}}$$

In summary: ω index is the standardized form of difference between the observed matches and expected matches of suspected copiers' and sources' responses.

$$\omega = \frac{\text{Observed Matches} - \text{Expected Matches}}{\text{Standard Error}}$$

Figure 1 shows the response probabilities of a given copier and source under Nominal Response Model of IRT for a multiple-choice test item. To compute ω statistic, response probabilities of each item is required. Figure 1 is used to explain the calculation of expected response probabilities.

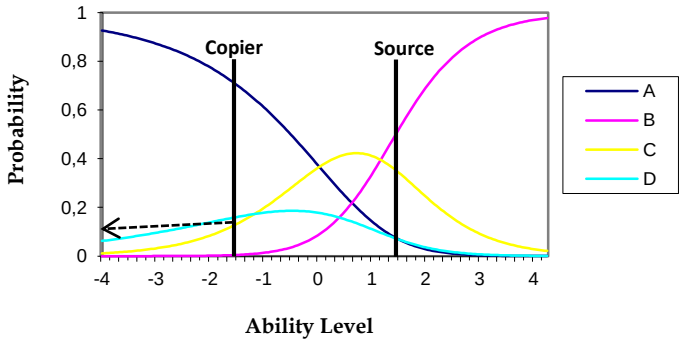


Figure 1. Response Probabilities of a Four-Option Multiple-Choice Item under Nominal Response Model

Let's say the suspected source has selected the option C. To obtain the expected value of the copier to select option C, a dashed line will be used.

A literature review for ω index is given below:

Wollack (1997) compared g_2 and ω indices in terms of answer copying type, sample size, test length, and amount of cheating (cheating ratio). The study showed that ω performed better than g_2 for all simulation conditions regarding Type I error rate and power.

Wollack and Cohen (1998) investigated the Type I error rates and power study of ω indices by using true and estimated ability parameters for sample sizes 100 and 500. The study showed that using estimated ability parameters instead of true ability parameters did not have a significant effect on Type I error rates. However, using estimated ability parameters instead of true ability parameters caused a slight decrease in power for the 100 sample size while remaining same power for the 500 sample size.

Sotaridona and Meijer (2002) compared K, $\overline{K}_1$, $\overline{K}_2$, and ω indices in terms of sample size, test length, and answer copying amount (ratio). The study showed that all indices performed well for Type I error rates. In addition, ω performed better than other

indices for power study. Another result of their study showed that, for cases which were not appropriate for ω, $\overline{K}_2$ performed better than K and $\overline{K}_1$ in terms of power.

Sotaridona and Meijer (2003) compared their S_1 and S_2 indices with K, $\overline{K}_2$, and ω in terms of sample size, test length, and amount of answer copying. Their study showed that S_1 was more powerful than $\overline{K}_2$. In addition, S_2 and ω indices were more powerful than other indices for answer-copying detection. In addition, with the appropriate estimate of item parameters from Nominal Response Model of Item Response Theory, ω performed better than other indices for any ability level of copier, and ω could be used for answer-copying detection for small sample sizes.

Wollack (2003) compared *Scrutiny!*, K , g_2, and ω. He showed that ω was the best performed index for all conditions.

Wollack (2006) suggested simultaneous use of $\overline{K}_2$, S_1, S_2, ω, H, and B indices. He showed that separate use of $\overline{K}_2$, S_1,S_2,ω,and B indices performed well for all nominal alpha levels (0.01, 0.005, 0.001, and 0.0005) in terms of Type I error rate; however, H index had higher Type I error rates. In addition, ω had best power study results among other indices. ω was followed by S_2 in terms of power rates. Simultaneous use of ω – H^* pair performed better than other index combinations. S_2 was stated as the most powerful index when it was not possible to compute ω index.

Sotaridona et al. (2006) showed that Kappa statistics had satisfactory results for five-option multiple-choice tests that have 30 and 60 test lengths. However, Kappa statistics for answer-copying detection was found to be sensitive to the ability levels of copier and source. Kappa statistics got high Type I error rates for 0.05 nominal alpha levels when the ability of copier was close to ability level of source.

Zopluoglu and Davenport (2012) compared Type I error rates and power of ω and GBT by manipulating ability levels of copier and source. The study showed that GBT index was slightly more powerful than ω index. In addition, it was observed that both indices were sensitive to the amount of cheating, and they couldn't detect low amounts of answer copying. The power did not reach 0.50 unless the copier examinee copied 50% of the answers of the source examinee whose ability levels were greater than 1.00, or unless the copier examinee copied 80% of the answers of the source examinee whose ability levels were greater than 2.00.

Zopluoglu (2016) investigated the performance of ω, GBT, K, $\overline{K}_1$, $\overline{K}_2$, S_1,andS_2 indices for simulated and real response data sets by using different Item Response Theory Models(one- [1PL], two- [2PL], three-parameter [3PL] models, Nominal Response Model [NRM]) by the area under the receiver operating characteristic curve (ROC). In addition, difficulty of items that were copied (random copying and difficulty-weighted copying) and test difficulty (easy, medium) was taken into account in the investigation. The results of the study showed that using NRM outcomes for 20% answer copying increased the performance. Another finding from the study was the slight differences between the performance of indices for 40% and 60% answer copying ratio. In the medium-difficulty test, slight increase of performance was

observed for all conditions for the difficulty-weighted copying. Furthermore, consistencies were observed between the results of real and simulated data sets.

Purpose of the Study

The literature review shows several studies that investigate the performance of ω and compare its performance with other indices in terms of Type I error and power studies. The most recent studies investigated performance of several indices by crossing the ability levels for several amounts of copying and several types of tests and items. Only one study accounts for the properties of items which were copied. In addition to ability levels and amount of answer copying, item parameters are effective for the performance of answer-copying indices and need to be investigated, because item and ability restrictions might affect the similarity of responses of examinees. In this study, the effects of item parameters (a and b parameters) on the ω index will be investigated.

Method

Data Generation

For this study, a five-option multiple-choice raw data set was generated by using GEN3PL_RawDATA_V2 (Luecht, 2011) for an 80-item test length. Generation was conducted for 10000 examinees with standard normal ability distribution *N(0~1)*. The scaling constant D was set at1.00. Options were A, B, C, D, and E for all items. Range of item parameters (a and b) was divided into two categories: the a parameter ranged between (0.10 – 0.80) for low-discriminative items and between (0.81 – 1.50) for high-discriminative items. The b parameter ranged between (-2.50 – 0.00) for easy items and between (0.01 – 2.50) for difficult items. By crossing item parameter ranges, a four-cell table was obtained. The table regarding item parameter ranges is given below.

Table 1

Item Parameters for Item Groups

Item Parameters		b	
		-2.50 – 0.00	0.01 – 2.50
a	0.10 – 0.80	Item Group 1	Item Group 2
	0.81 – 1.50	Item Group 3	Item Group 4

20 items were generated for each item group cell and 4x20=80 items for the total test. Item Group 1 contains easy and low-discriminative items, Item Group 2 contains difficult and low-discriminative items, Item Group 3 contains easy and high-

discriminative items, and Item Group 4 contains difficult and high-discriminative items.

Procedure

To evaluate the ability level effects, ability range of examinees was divided into four categories (-3.00 – -1.50, -1.50 – 0.00, 0.00 – 1.50 and 1.5 – 3.00). By crossing ability levels for copier and source, a sixteen-cell table was obtained.

After crossing ability levels, sixteen different copier-and-source pair types were obtained. No cheating data was used for Type I error study. Cheating scenarios were implemented for each item group for the power study with 100 replications for three nominal alpha levels (0.05, 0.01, and 0.001). Amount of copying was held constant for the whole study. For the Type I error study, 16 different conditions were provided by ability crossing (4x4=16). With 100 replications, 16x100= 1600 honest pairs were examined by ω index. For the power study, simulation conditions are given in Table 2.

By crossing all conditions, 2x2x16x1x100=6400 pairs were investigated by ω index. Item parameter generation for given restrictions, ω index calculations, data and output management procedures were conducted by using R (2016) programming language. Graphs were obtained from Excel.

Table 2

Simulation Conditions for Power Study

Condition	Number of Levels	Level Values
Test Length	1	80
Item Difficulty (b parameter)	2	-2.50 – 0.00 0.01 – 2.50
Item Discrimination (a parameter)	2	0.10 – 0.80 0.81 – 1.50
Ability Range (Source and Copier)	4	-3.00 – -1.50 -1.50 – 0.00 0.00 – 1.50 1.50 – 3.00
Amount of Copying	1	25%
Number of Replications	100	

Results

Type I Error Study

The results of the Type I error study are given in Figure 2. When we investigated Figure 2, it was observed that Type I error rates of ω index were generally close to or below related nominal alpha levels, with some exceptions.

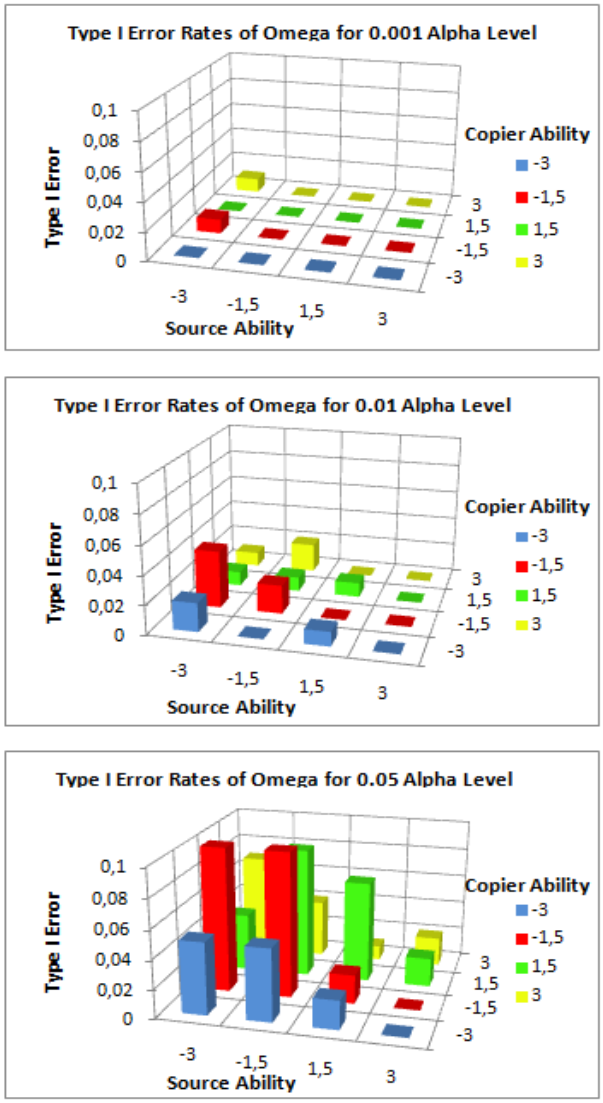


Figure 2. Results of Type I Error Study for 0.001, 0.01, and 0.05 Nominal Alpha Levels

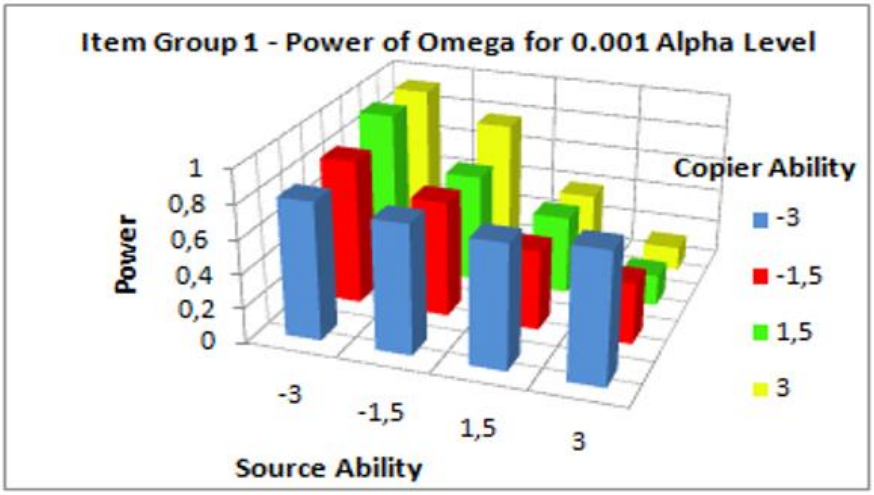
Type I error rates ranged between 0.11 and 0.00. 0.11 error rates were observed when -1.50 – 0.00 ability range copiers cheated from source whose abilities were below 0.00. It shows that the ability range of the source is an important factor for Type I error rates, and lower source ability levels may inflate Type I error rates. Type I error rates for higher source ability levels seemed quite satisfactory, even with the high copier ability levels for all nominal alpha levels.

Power Study

The results of the power study are given below, separately for each item group.

Results for Item Group 1: Low-Discriminative and Easy Items

The results of the power study for item group 1 (low-discriminative and easy items)are given in Figure 3. When we observe Figure 3, it can be seen that power rates of the ω index range between 1.00 and 0.14. To summarize general tendency, power rates tend to decrease from 0.05 to 0.001 nominal alpha levels. Particularly for the 0.001 nominal alpha level, most of the results seem unsatisfactory in terms of power. An increase of power rate was observed to the low source ability levels, and best power rates for item group 1 are observed when the ability of the source ranged between -3.00 and -1.50. Another inference might be the decrease of copier ability levels, where the power rates tend to increase especially for 0.05 and 0.01 nominal alpha levels.



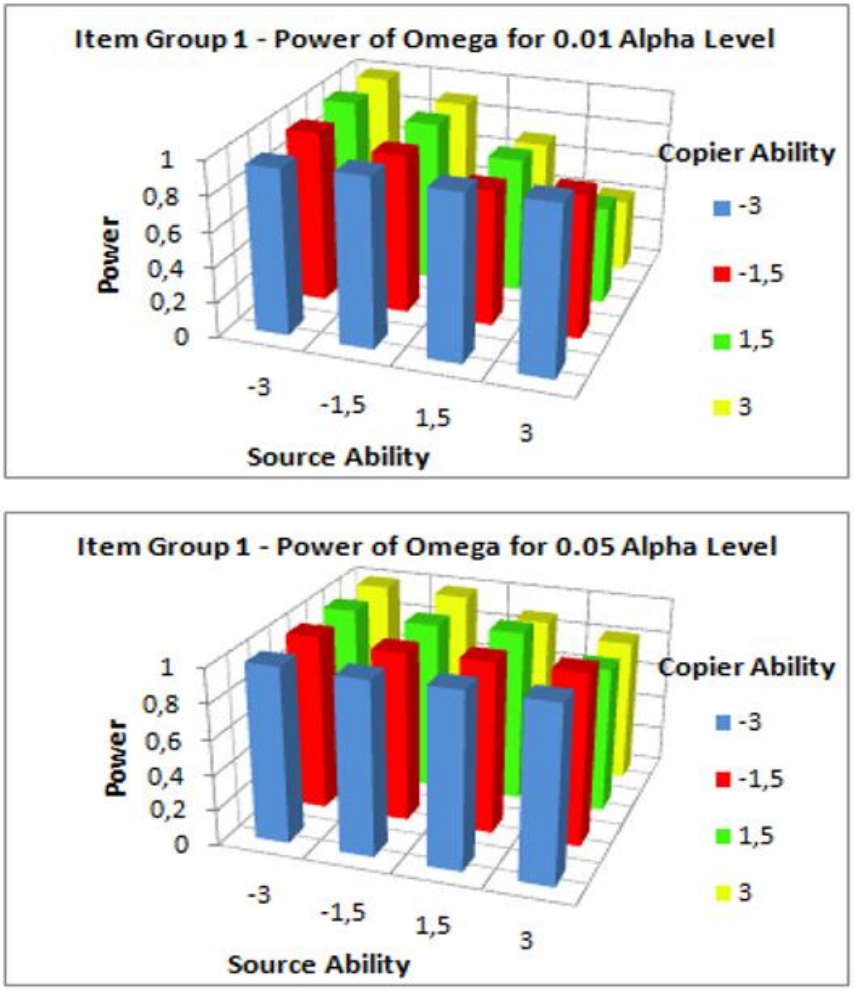


Figure 3. Results for Item Group 1 for 0.001, 0.01, and 0.05 Nominal Alpha Levels

Results for Item Group 2: Low-Discriminative and Difficult Items

The results of the power study for item group 2 (low-discriminative and difficult items) are given in Figure 4. When we observe Figure 4, it can be seen that power rates of the ω index ranged between 0.93 and 1.00. Nearly all power rates for all copier and source ability combinations were very close to 1.00. ω index performed very satisfactory for item group 2, despite the low-discriminations item statistics.

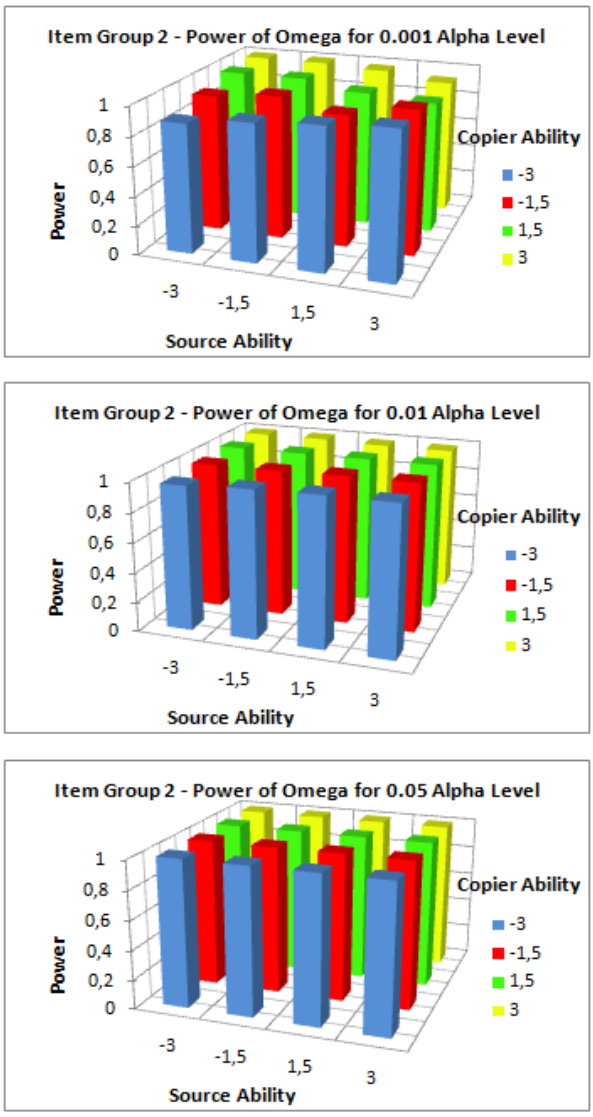


Figure 4. Results for Item Group 2 for 0.001, 0.01, and 0.05 Nominal Alpha Levels

Results for Item Group 3: High-Discriminative and Easy Items

The results of the power study for item group 3 (high-discriminative and easy items) are given in Figure 5.

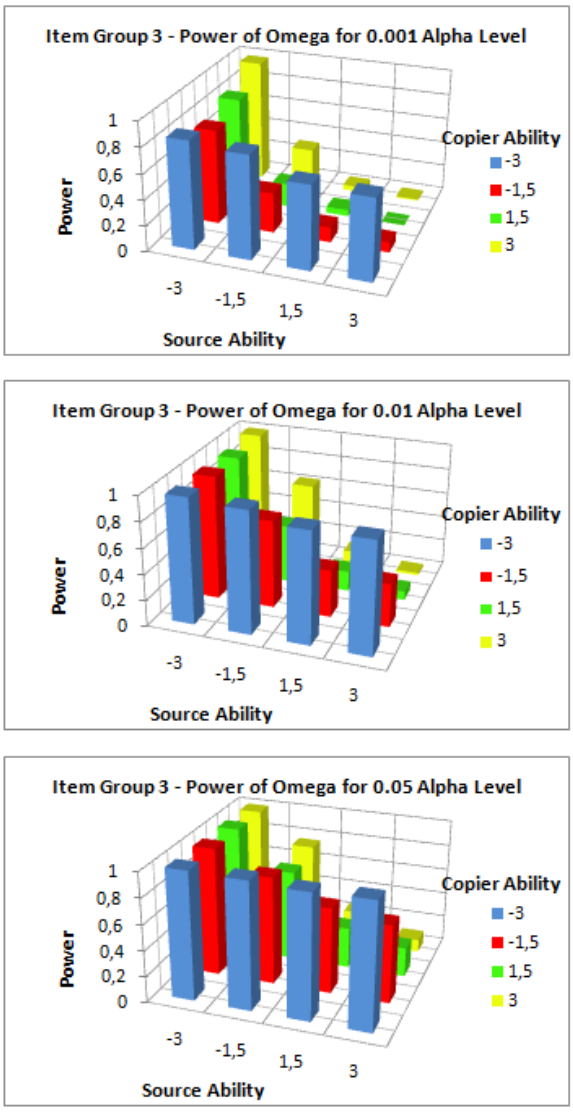


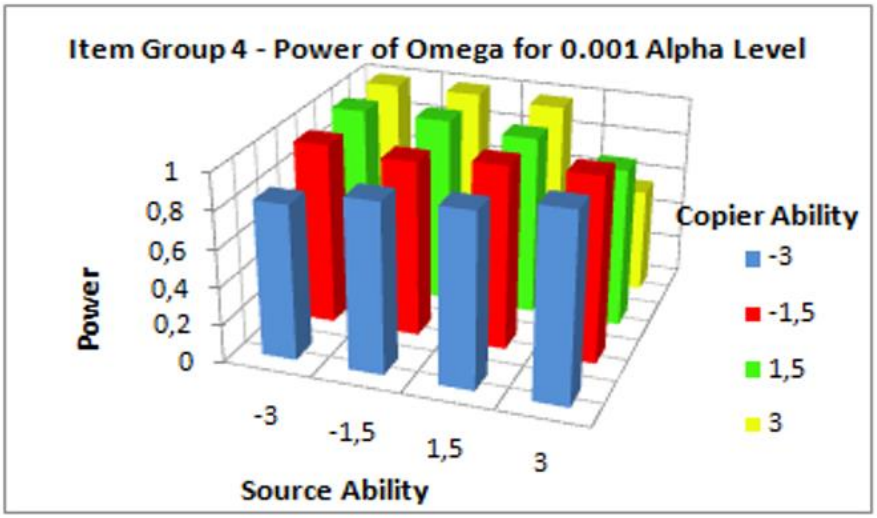
Figure 5. Results for Item Group 3 for 0.001, 0.01, and 0.05 Nominal Alpha Levels

When we observe Figure 5, it can be seen that power rates of the ω index ranged between 1.00 and 0.00. ω showed extremely bad performance for several condition cells. To summarize the general tendency, power rates tend to increase with the

decrease of copier ability levels and of source ability levels for all nominal alpha levels for item group 3. The best power rate results of the ω index were obtained when the ability range of copier and source were between -3.00 and 1.50. Other results for other ability combinations seem unsatisfactory for all nominal alpha levels. The only exception maybe the results for the conditions in which copiers' ability levels ranged between -3.00 and -1.50 for 0.05 nominal alpha level.

Results for Item Group 4: High-Discriminative and Difficult Items

The results of the power study for item group 4 (high-discriminative and difficult items) are given in Figure 6. When we observe Figure 6, it can be seen that power rates of the ω index ranged between 0.56 and 1.00. Most power rates for all copier and source ability combinations were very close to 1.00. ω index performed very satisfactorily for item group 4 between all the conditions of this research.



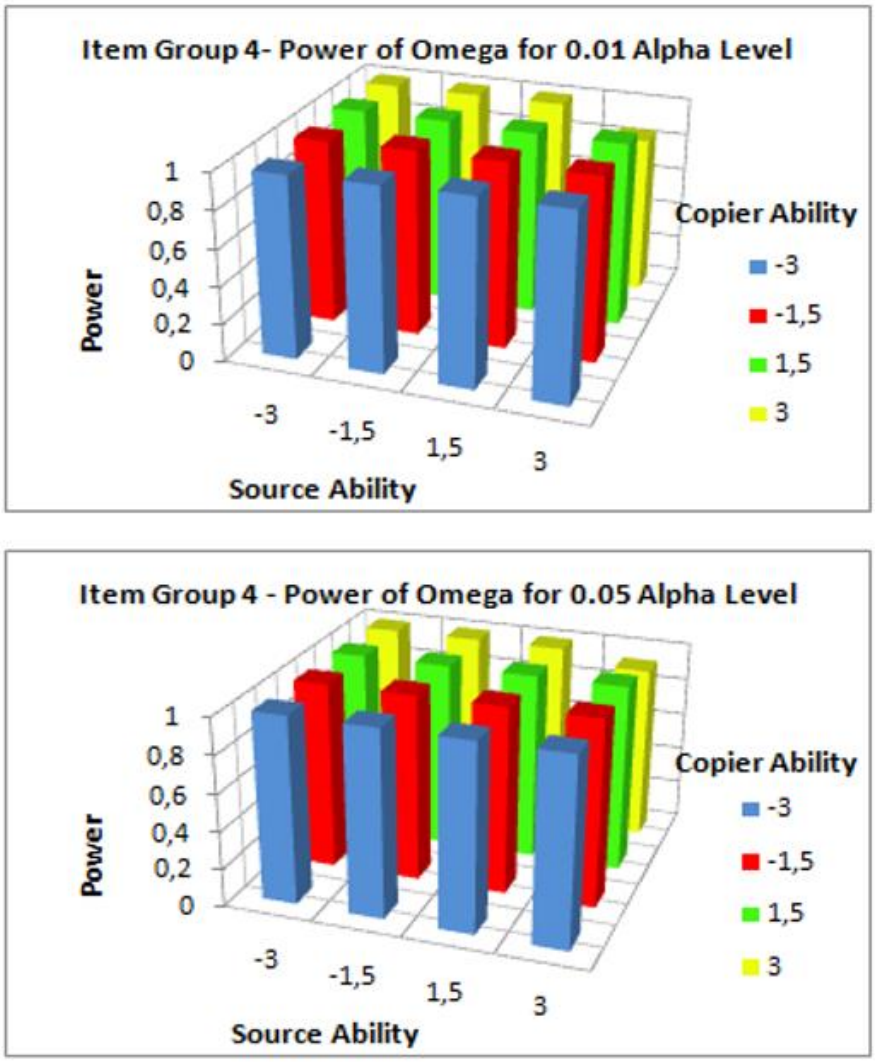


Figure 6. Results for Item Group 4 for 0.001, 0.01, and 0.05 Nominal Alpha Levels

Discussion and Conclusion

Type I error rates and power studies are vitally important in evaluating the importance of statistics or indices for decision making. ω index is one of the most popular indices for detecting answer copying. Several studies investigate and compare the performance of ω statistic with other answer-copying detection statistics. The literature review shows that test length, amount of copying, IRT model used for item parameter estimation, and ability levels of copying are effective for the performance

of the ω index. In this study, the amount of copying (25%) and the test length (80 items) were held constant, with focus on the interaction of copied item properties in term of item parameters as well as copier and source ability level crossings. Results of the Type I error study showed that the ω index performed well for nearly all research conditions for all nominal alpha levels with some exceptions. The highest Type I error rate 0.11 was obtained when -1.50 – 0.00 ability range copiers cheated from sources whose abilities were below 0.00.

Researchers should be careful when using the ω index to compare suspected copiers with low ability sources in terms of false positives. When we evaluated the power study results, it could be seen that item difficulty was very effective for the power study results. If the copiers copied from difficult items, the power of ω statistic accelerated immediately, and most of them were acceptable. However, copying the answers of easy items decreased the power immediately, and most of them were unacceptable. Best results for easy item cheaters were found when copiers' and sources' ability levels were between -3.50 and -1.50. Since easy items are more informative for low ability ranges, item information might cause the increase of power for low ability levels. When we evaluate the effect of the discrimination parameter, it was not as dominant as item difficulty. However, best results were obtained when copiers cheated from discriminative and hard items, and the worst results were obtained from discriminative and easy items. More research need to be conducted that integrates the different amounts of copying with test and item-information functions.

References

Angoff, W.H. (1974). The development of statistical indices for detecting cheaters. *Journal of American Statistical Association, 69*, 44-49.

Anikeef, A.M. (1954). Index of collaboration for test administrators. *Journal of Applied Psychology, 38*, 174-177.

Armstrong, R. D., & Shi, M. (2009). A parametric cumulative sum statistic for person fit. *Applied Psychological Measurement*, *33*(5), 391-410.

Assessment Systems Corporation (1993). *Scrutiny!: Software to identify test misconduct.* Advanced Psychometrics.

Bay, M. L. G. (1994). Detection of copying on multiple-choice examinations (Doctoral dissertation, Southern Illinois University, 1987). *Dissertation Abstracts International, 56(3-A)*, 899.

Bellezza, F.S., & Bellezza, S.F. (1989). Detection of cheating on multiple-choice tests by using error-similarity analysis. *Teaching of Psychology, 16*, 151-155. *British Journal of Arts and Social Sciences ISSN: 2046-9578* 59.

Belov, D. I. (2011). Detection of answer copying based on the structure of a high-stakes test. *Applied Psychological Measurement*, 35(7), 495–517.

Belov, D. I., & Armstrong, R. D. (2010). Automatic detection of answer copying via kullback-leibler divergence and K-index. *Applied Psychological Measurement*, 34(6), 379–392.

Bird, C. (1927). The detection of cheating in objective examinations. *School and society, 25*, 261-262.

Bird, C. (1929). An improved method of detection cheating in objective examinations. *Journal of Educational Research, 25*, 261-262.

Cizek, G. J., & Wollack, J. A. (2017). *Handbook of quantitative methods for detecting cheating on tests*. New York, NY: Routledge.

Clark, J. M. (2010). *Aberrant response patterns as a multidimensional phenomenon: Using factor-analytic model comparison to detect cheating*. *ProQuest LLC*. University of Kansas.

Deng, W. (2008). An innovative use of the standardized log-likelihood statistic to evaluate person fit. Dissertation Abstracts International Section A: Humanities and Social Sciences. Rutgers State University of New Jersey.

Frary, R. B. (1993). Statistical detection of multiple-choice answer copying: Review and commentary. *Applied Measurement in Education, 6*, 153-65.

Frary, R. B., Tideman, T. N., & Watts, T. M. (1977). Indices of cheating on multiple-choice tests. *Journal of Educational Statistics, 6*, 152-165.

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.

Harpp, D.N., Hogan, J.J., & Jennings, J.S. (1996). Crime in the classroom – Part II, an update. *Journal of Chemical Education*, 73(4), 349-351.

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-94-4). Princeton, NJ: Educational Testing Service.

Hui, H.-fai. (2010). *Stability and sensitivity of a model-based person-fit index in detecting item pre-knowledge in computerized adaptive test. Dissertation Abstracts International Section A: Humanities and Social Sciences*. University of Hong Kong.

Luecht, R. M. (2011). *Gen3PL Raw Data (Version 2)*. Greensboro, NC: [Author].

Maynes, D. D. (2009, April). *Combining statistical evidence for increased power in detecting cheating*. Presented at the annual conference of the National Council on Measurement in Education, San Diego, CA.

Saupe, J.L. (1960). An empirical model for the corroboration of suspected cheating on multiple-choice tests. *Educational and Psychological Measurement, 20*, 475-489.

Shu, Z. (2011). *Detecting test cheating using a deterministic, gated item response theory model*. (Doctoral dissertation, The University of North Carolina at Greensboro, 2010). *Dissertation Abstracts International Section A: Humanities and Social Sciences*.

Sotaridona, L.S., & Meijer, R.R. (2002). Statistical properties of the K-index for detecting answer copying. *Journal of Educational Measurement*, *39*, 115-132.

Sotaridona, L. S., & Meijer, R. R. (2003). Two new statistics to detect answer copying. *Journal of Educational Measurement, 40*, 53-69.

Sotaridona, L.S., van der Linden, W.J., & Meijer, R.R. (2006). Detecting answer copying using the kappa statistic. *Applied Psychological Measurement, 30*, 412-431.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

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*, 361-378.

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.

van Krimpen-Stoop, E. M. L. A., & Meijer, R. R. (2001). CUSUM-based person-fit statistics for adaptive testing. *Journal of Educational and Behavioral Statistics*, *26*(2), 199–217.

Wesolowsky, G. O. (2000). Detecting excessive similarity in answers on multiple choice exams. *Journal of Applied Statistics, 27(7),* 909-921.

Wollack, J. A. (1997). A nominal response model approach to detect answer copying. *Applied Psychological Measurement, 21*, 307-320.

Wollack, J. A. (2003). Comparison of answer copying indices with real data. *Journal of Educational Measurement*, *40*, 189-205.

Wollack, J. A. (2006). Simultaneous use of multiple answer copying indexes to improve detection rates. *Applied Measurement in Education, 19*, 265-288.

Wollack, J. A., & Cohen, A. S. (1998). Detection of answer copying with unknown item and trait parameters. Applied Psychological Measurement, 22, 144-152.

Wollack, J. A., & Maynes, D. D. (2017). Detection of test collusion using cluster analysis. In G. J. Cizek and J. A. Wollack (Eds.), *Handbook of quantitative methods for detecting cheating on tests* (pp. 124-150). New York, NY: Routledge.

Zopluoglu, C. (2016). Classification performance of answer-copying indices under different types of irt models. *Applied Psychological Measurement, 40* (8), 592-607.

Zopluoglu, C., & Davenport, E.C., Jr. (2012). The empirical power and type I error rates of the GBT and ω indices in detecting answer copying on multiple-choice tests. *Educational and Psychological Measurement*, *1-26*.

Madde Parametreleri ve Yetenek Düzeylerine göre Omega İndeksinin Performansının İncelenmesi

Atıf:

Sunbul O. & Yormaz S. (2018). Investigating the performance of omega index according to item parameters and ability levels, *Eurasian Journal of Educational Research*, 74, 207-226, DOI: 10.14689/ejer.2018.74.11

Özet

Problem Durumu: Çoktan seçmeli maddeler sağlamış olduğu birtakım avantajlardan dolayı geniş ölçekli sınavlarda sıkça kullanılmaktadır. Bütün sınavlarda olduğu gibi çoktan seçmeli maddelerden oluşan testlerin de geçerlik ve güvenirliği bireyler hakkında karar vermede oldukça önemli bir role sahiptir. Sınav süreçlerinde bireylerin çeşitli şekillerde kopya çekme davranışı gösterme eğiliminde olduğu bilinmektedir. 21. yüzyılda teknoloji giderek ivmelenen bir şekilde gelişim göstermiştir. Teknolojide meydana gelen bu gelişmelerin bireylerin yaratıcılığı ile birleşmesi ve bireyler arasındaki rekabetin artması sonucu çok çeşitli kopya çekme türleri ortaya çıkmıştır. Testin uygulanması öncesinde, test uygulaması sürecinde ve sonrasında oluşabilecek hile karıştırma girişimlerini engellemek ağırlıklı olarak test güvenliğini ilgilendiren süreçlerdir. İstenmeyen bir durum oluşmaması adına test güvenliği üzerinde hassasiyetle durulmalıdır. Eğer gerekli hassasiyet gösterilmezse, doğacak sonuçların maliyeti ağır olabilmektedir. Bu hassasiyete rağmen bir kopya durumu oluştuğu takdirde, kopyayı tespit edebilmek için alternatif yol arayışlarına girilmiştir ve birçok kopya belirleme yöntemi geliştirilmiştir. Bu yöntemlerin en çok itibar edilenleri istatistiksel yöntemlerdir. İstatistiksel yöntemlerin performansına yönelik yapılan araştırmalarda ise madde tepki kuramına dayalı olan ω indeksinin kopya belirlemede öne çıktığı görülmüştür. Alanyazında ω indeksinin hangi koşullar altında çalıştığını incelemek üzere çeşitli araştırmalar bulunmaktadır. Aynı zamanda I. Tip hata oranı ve gücü bu koşullar altında diğer indekslerle karşılaştırılmıştır. Bunlardan çoğu çeşitli örneklem büyüklüğü, test maddesi ve kopya oranı ile bireylerin yetenek düzeyleri çaprazlanarak ω indeksinin performansı ortaya konmaya çalışılmıştır.

Araştırmanın Amacı: Bu çalışmanın amacı alanyazında henüz ortaya konmayan ω indeksinin madde güçlük, madde ayırt edicilik ve yetenek sınırlandırmaları altındaki performansını incelemektir.

Araştırmanın Yöntemi: Çalışmada beş seçenekli çoktan seçmeli 80 maddeden oluşan veri seti GEN3PL_Raw DATA_V2 yardımıyla elde edilmiştir. Veri standart normal yetenek dağılımına *N(0~1)* sahip 10000 birey için üretilmiştir. Ölçekleme sabiti olan D ise 1.00 olarak alınmıştır. a ve b madde parametreleri iki kategoriye ayrılarak incelenmiştir. a parametresinde düşük ayırt edici maddeler için (0.10 – 0.80) aralıkları, yüksek ayırt edici maddeler için (0.81 – 1.50) aralıkları ele alınmıştır. b parametresinde kolay maddeler için (-2.50 – 0.00) aralıkları, zor maddeler için (0.01 – 2.50) aralıkları ele alınmıştır. Parametre aralıkları çaprazlanarak dört hücreli tablo elde edilmiş ve böylece dört farklı madde grubu ortaya çıkmıştır. Her bir madde grubu için 20 maddelik veri üreterek toplamda 4x20=80 maddelik veri seti elde edilmiştir. 1. Madde Grubu kolay ve düşük ayırt edici maddelerden, 2. Madde Grubu zor ve düşük ayırt edici maddelerden, 3. Madde Grubu kolay ve yüksek ayırt edici maddelerden ve 4. Madde Grubu zor ve yüksek ayırt edici maddelerden oluşmaktadır.

Kaynak ve kopyacı bireyin yeteneğinin, ω indeksinin performansı üzerine etkisini incelemek için (-3.00 – -1.50, -1.50 – 0.00, 0.00 – 1.50, 1.50 – 3.00) olmak üzere dört kategoriye ayrılmıştır. Kopyacı ve kaynak çifti için yapılan çaprazlamalar sonucu oluşan 16 hücre için kopya durumu oluşturup her bir veri için indeksin gücü incelenmiştir. I. Tip hata için ise kopya durumu oluşturulmadan inceleme yapılmıştır. İşlemler için 100 replikasyon yapılmıştır. Her replikasyondan elde edilen sonuçlar kullanılarak belirlenen α düzeylerine (0.001, 0.01 ve 0.05) göre indeksin I. Tip hata oranları ve kopya belirleme güçleri hesaplanmıştır. I. Tip hata çalışmasında 16 farklı koşulla 100 replikasyon sonucunda 16x1000= 16000 kopya çekmeyen birey çifti için ω indeksine ait çıktılar elde edilmiştir. Güç çalışmasında ise sabit kopya oranı (%25) ile 2x2x16x1x100=6400 birey çifti için çıktılar elde edilmiştir. Madde parametrelerinin üretimi, ω indeksinin hesaplanması ve veri ve çıktıların elde edilmesinde R programlama dili kullanılmıştır.

Araştırmanın Bulguları: I. Tip hata çalışmasında elde edilen bulgulara göre ω indeksinin I. Tip hata oranı ilgili alfa düzeyine genellikle yakın ya da altında değerler almıştır. Bu değerler 0.11 ile 0.00 arasında değişmektedir. I. Tip hata oranının 0.11 olduğu durumlar, -1.50 – 0.00 yetenek aralığında yer alan kopyacının 0.00 altındaki yetenek düzeyine sahip kaynaktan çektiği kopya durumlarında gözlenmiştir.

Güç çalışması sonucunda kolay ve düşük ayırt edici maddelerden oluşan 1. madde grubundan elde edilen bulgulara göre indeksin gücü 1.00 ve 0.14 arasında değişmektedir. Alfa düzeyi 0.05'ten 0.001'e düştükçe güçte azalma eğiliminin olduğu görülmektedir. Düşük yetenek düzeyine sahip kaynağın yer aldığı kopya durumlarında güç bir miktar artmakta ve en yüksek gücün kaynağın yetenek düzeyinin 3.00 ve -1.50 aralığında olduğu durumlarda gözlenmiştir. 0.05 ve 0.01 alfa düzeyinde kopyacının yetenek düzeyinin düştüğü durumlarda güçte artışın olduğu görülmektedir.

Madde grubu 2'de ise düşük ayırt edici maddeler olmasına rağmen tüm kopyacı ve kaynağın yetenek düzeylerinin çaprazlanmasında ω indeksinin gücünün 1.00'e yakın olduğu ortaya çıkmıştır.

Kolay ve yüksek ayırt edici maddelerin yer aldığı madde grubunda ise indeksin gücünün genelde oldukça düşük olduğu görülmüştür. Kaynağın yetenek düzeyinin arttığı ve kopyacının yetenek düzeyinin azaldığı kopya durumlarında gücün artma eğiliminde olduğu ve kopyacı ve kaynak çiftinin yetenek düzeyi -3.00 ile 1.50 aralığında iken gücün bu grupta en yüksek değerler aldığı ortaya çıkmıştır. Diğer koşullarda ise sadece 0.05 alfa düzeyinde -3.00 ve -1.50 yetenek düzeyi aralığı haricinde tüm alfa düzeyinde ve tüm koşullarda ω indeksinin performansının düşük olduğu görülmüştür.

Zor ve yüksek ayırt edici maddelerin kopya çekildiği durumlarda ise ω indeksinin gücü 0.56 ile 1.00 arasında değişmektedir. Birey çiftlerinin yetenek düzeylerine ait kombinasyonların tümünde çoğunlukla gücün 1.00'e yakın olduğu ortaya çıkmıştır.

Araştırmanın Sonuçları ve Önerileri: ω indeksinin I. Tip hata oranı bazı durumlar haricinde tüm koşullar ve tüm alfa düzeylerinde birbirine yakın ve düşük değerler almıştır. Ancak özellikle kopyacının yetenek düzeyinin düşük olduğu durumlarda I. Tip hata oranında fazla artışın olduğu görülmektedir. Araştırmacılar düşük yetenek düzeyine sahip kaynağın yer aldığı kopya durumlarında ω indeksini kullanırken dikkatli olmalıdırlar. Güç çalışması sonucunda ise madde güçlüğünün indeksin gücünü oldukça etkilediği görülmektedir. Kolay maddelerde çekilen kopyayı belirlemede indeksin gücünde anlamlı bir azalmanın olduğu ortaya konmuştur. Kolay maddeden kopyanın çekildiği durumlarda en iyi sonuçlar kopyacı ve kaynağın yetenek düzeyinin -3.50 ile -1.50 aralığında olduğunda ortaya çıkmıştır. Madde ayırt ediciliğin ise güçlük kadar indeksin gücünde etkili olmadığı ortaya çıkmıştır. Ancak ω indeksi, ayırt edici ve zor maddelerden çekilen kopyayı belirlemede en güçlü, ayırt edici ve kolay maddelerden çekilen kopyayı belirlemede ise en zayıf olduğu görülmüştür.

Anahtar Kelimeler: Cevap kopyalamayı belirleme, hile, test güvenliği.