ROCZNIKI PSYCHOLOGICZNE / ANNALS OF PSYCHOLOGY
2014, XVII, 2, 457-473

KRZYSZTOF FRONCZYK
The University of Finance and Management in Warsaw
Faculty of Psychology

THE IDENTIFICATION OF RANDOM OR CARELESS RESPONDING IN QUESTIONNAIRES: THE EXAMPLE OF THE NEO-FFI

The article presents two little-known indices of random or careless responding: Cattell's sabotage index and fixed individualized chance (FIC) score. Both indices are used to identify people who provide content-irrelevant answers, such as random ones, in multidimensional questionnaires. The aim of the study was to verify empirically the diagnostic applications of these indices in distinguishing the actual NEO-FFI scores from random data generated by a computer. The study involved 943 participants and 1000 randomly generated protocols. Based on both indices in combination and using logistic regression, it proved possible to distinguish the actual data from the random data fairly well. Approximately 86% of all data was classified correctly. This result is quite high, given that some participants might have responded to the NEO-FFI items in a random way.

Keywords: random responding, careless responding, Cattell's sabotage index, FIC index, NEO-FFI.

There are many factors that distort the results obtained in questionnaire-based personality studies. Nichols, Greene, and Schmolck (1989) distinguished two basic types of response distortions. The first one is *content-responsive faking*. Responses affected by this type of distortion are content-relevant but do not reflect the respondents' actual self-knowledge; examples include simulating or

Corresponding author: KRZYSZTOF FRONCZYK – Faculty of Psychology, The University of Finance and Management in Warsaw, ul. Pawia 55, 01-030 Warszawa; e-mail: fronczyk@vizja.pl
The research was partly financed with grant no. 2012/07/B/HS6/00933 awarded by the Polish National Science Center (NCN).

minimizing certain symptoms or answering in accordance with social expectations ("fake bad" or "fake good"). This is a kind of either situation-dependent or personality-dependent self-presentation (Zawadzki, 2006). This paper is devoted to the other type of distortions distinguished by Nichols and colleagues (1989), namely the distortions unrelated to the content of the test items: *content nonresponsivity* or *noncontent responding*. Distortions of this kind include leaving questions unanswered as well as certain responding styles, such as random or careless responding (Beach, 1989). In Cattell's terminology, this manner of responding can be called sabotage (Cattell, Eber, & Tatsuoka, 1970).

Careless responding is usually understood as giving inconsistent responses to questions with similar content or to pairs of questions the answers of which are correlated. Random responding is defined as responding without regard to the content of test items, as if the participants randomly indicated their answers to the questions without reading them (Evans & Dinning, 1983). Moreover, there is no consistency in random responding, which means that distortions of this kind cannot be interpreted as manifesting a certain way of self-presentation (Nichols & Greene, 1997). Both conceptually and empirically, distinguishing between random and careless responding is quite difficult. Most of the existing indices are comparably sensitive to those two distortion types. In this paper, random and careless responding will therefore be treated jointly.

Random answers increase error variance, which decreases the intercorrelation of test items leading to a decrease in the questionnaire's internal consistency. Similarly, on the individual level, the more random answers a particular person has given, the less consistent his or her answers are. In this paper, two out of a number of indices used for diagnosing random responding will be discussed in more detail. These are the sabotage index developed by Cattell, Eber, and Tatsuoka (1970) and fixed individualized chance score (Haertzen & Ross, 1978). Each of these indices was constructed in a different way.

Cattell's Sabotage Index

The sabotage index used in Cattell's Sixteen Personality Factor Questionnaire (16PF) was developed on the assumption that each questionnaire scale may be divided into two parts (Cattell et al., 1970). The score computed on the basis of one half of the questions making up a given scale can be estimated, using the regression method, on the basis of the other half of the questions making up the same scale. Such estimation is possible if we assume that participants respond by referring to their self-knowledge, that they do not respond randomly, and that they do not follow any responding styles. Thus, based on empirical data, it is possible to compute two groups of scores: the actual scores for each of the two parts of a given scale and the scores estimated on the basis of the other part. The combined sabotage index is computed as the sum of squared differences between the actual scores for a half of each questionnaire scale and the scores estimated based on a linear regression of the first-half score on the second-half score. In other words, this index is the sum of squared residuals from the regression of each scale's first-part score on the second-part score computed for each scale of the questionnaire. The lower the value of this index, the lower the probability that the person for whom the index was computed responded in a random or careless manner. This is because a low value indicates a fairly high concordance between the distinguished parts of individual scales. Practically, Cattell and colleagues (1970) recommended that the sabotage index should be computed using the averaged correlation between pairs of half-scales and that the calculations should be carried out on normalized scores. This was probably meant to simplify the calculations.

In the case of Cattell's Personality Questionnaire, the sabotage index did not bring particularly good results with regard to the identification of random answers (O'Dell, 1971). Some scores were classified as random even though they were the actual scores obtained by participants. Better results were obtained using other control scales (Irvine & Gendreau, 1974). However, it is possible that for other questionnaires, with scales having higher internal consistency, this index would bring better results.

It is also possible that the result obtained by O'Dell (1971) is a by-product of the simplified procedure of sabotage index computing. It may be assumed that the applicability of the discussed index should be better if the exact values of regression coefficients for the halves of each scale of a questionnaire were used (instead of averaged values for all the scales). Since no other findings concerning this issue are available, apart from the two analyses mentioned above (O'Dell, 1971; Irvine & Gendreau, 1974), this line of research is worth continuing.

Fixed Individualized Chance Score

A different approach to diagnosing random or careless responding is also possible. In this approach, the expected score on each scale of a questionnaire is determined, with the assumption that the participant responds randomly. If, at the individual level, the proportion of a given type of responses (e.g., "yes" responses) in random responding is the same in the whole questionnaire as well as in each of its scales, the fixed individualized chance score (FIC) can be computed for each scale based on the proportions of particular kinds of responses in the whole questionnaire, regardless of which scale they belong to in terms of content. This approach was applied by Haertzen and Ross (1978). In the simplest case, when a questionnaire with a two-category response format (“yes” – “no”) is considered, the expected score for a particular scale can be computed as follows:

$$FIC_i = (P_T \times T_i) + (Q \times N_i)$$

where:

FIC_i – fixed individualized chance score on the *i*th scale

P_T – the proportion of the person’s positive responses in the whole questionnaire, regardless of scale

T_i – the number of positive diagnostic responses in the *i*th scale

$Q = 1 - P_T$ – the proportion of the person’s negative responses in the whole questionnaire, regardless of scale

N_i – the number of negative diagnostic responses in the *i*th scale

The fixed individualized chance scores assuming random responding obtained for each scale can be compared with the actual scores. The differences between the fixed individualized chance score and the actual score for each scale, squared and added up, give the overall fixed individualized chance score for all the scales. The higher the value of this index, the lower the probability that the person responded in a random manner, since his or her scores diverge considerably from the profile determined on the basis of fixed individualized chance scores.

Just like in the case of Cattell’s sabotage index, publications on the fixed individualized chance score are few. Admittedly, this index has already been described in the Polish scientific literature (Paluchowski, 1983), but the studies on its validity are limited to the analyses carried out by Haertzen and Ross (1978) as well as by Ross and Haertzen (1979). Those researchers demonstrated a very high validity of their index using a little-known measure, the Social Experience Questionnaire. Two kinds of data were analyzed: data collected from actual respondents and randomly generated data. The index discussed allowed correct identification of all randomly generated data, and only 3.7% of the actual data were misclassified as random. However, this promising result was obtained using very special data from a limited number of respondents. No other studies using the fixed individualized chance score index are known. Therefore, further research must be carried out with larger groups and with more commonly used questionnaires.

The Methodology of Verifying the Validity of Random Responding Indices

In order to verify the diagnostic potential of various methods of identifying random or careless responding, simulation studies with random responses generated by a computer are often performed. They are used as criterion data juxtaposed with the actual data for various random responding indices (e.g., Archer & Elkins, 1999; Pinsoneault, 2002, 2005). The high validity of a given index is confirmed by its capacity to differentiate between actual and simulated data, which can be checked, for example, using discrimination analysis or regression analysis, as in the study by Baer, Ballenger, Berry, and Wetter (1997) or by Archer and Elkins (1999).

Simulation data are sometimes criticized as being of a different character than data obtained in real conditions. Moreover, despite responding in accordance with self-knowledge, some people may give answers that are untypical, as if they were random. Still, in research practice, computer simulation is the simplest method of obtaining data that are known to have been generated in a random way. Naturally, participants may be asked to respond in a random manner or without reading the contents of items. But such data could be similarly criticized as generated in an artificial situation and as detached from the real assessment conditions.

Research Objective and Assumptions

The aim of the study was to verify the validity and to select the better of the two indices of random or careless responding in the NEO-FFI questionnaire: the sabotage index and the fixed individualized chance score, as well as the combined use of these two indices. This particular questionnaire was chosen due to its considerable popularity in psychological practice and due to the fact that, unlike many other psychological inventories, it has no control scales or other responding style indicators.

It was decided that the indices used would be the sabotage index proposed by Cattell, Eber, and Tatsuoka (1970) and the fixed individualized chance score index proposed by Haertzen and Ross (1978); they were modified by the author for the purposes of the present study. Contrary to what Cattell originally proposed, the sabotage index was computed using raw data instead of normalized data. Additionally, exact regression coefficients were determined for each pair of half-scales and average coefficients were not used. Each NEO-FFI scale was divided into two halves, one of them comprising even-number items and the other comprising odd-number ones. When computing the sabotage index, regression was performed – for each of the five NEO-FFI scales separately – with the odd-number part of the scale as the dependent variable and the even-number as the independent variable.

The fixed individualized chance score index, discussed earlier, applied only to items with binary answers. To be fit for use in the present study, it was extended for use with measures consisting of items with a multicategory response format (modified fixed individualized chance score, MFIC). To obtain that, the individual proportions of each response category in the questionnaire as a whole had to be computed, and then those proportions were multiplied by the number of directly scored (positively keyed) items and, separately, by the number of reverse-scored (negatively keyed) items. In the next step, each of the products obtained as described above was multiplied by the number of points assigned in a given way of scoring. The final result is obtained by adding up all the products computed for individual categories as well as for both types of scoring. This is briefly presented by the following formula:

$$MFIC_i = \sum_{j=1}^{k} w_j P_j T_{ij} + \sum_{j=1}^{k} z_j P_j N_{ij}$$

where:

$MFIC_i$ – modified fixed individualized chance score on the ith scale

w_j – the number of points assigned to a directly scored category j response

z_j – the number of points assigned to a reverse-scored category j response

P_j – the proportion of category j responses given by a particular person in the whole questionnaire, regardless of scale

T_{ij} – the number of category j responses in the ith scale, with the w number of points assigned to this category

N_{ij} – the number of category j responses in the ith scale with the z number of points assigned to this category

In order to test the validity of the indices, empirical data collected in neutral conditions and randomly generated criterion data were used. It was assumed that random responses to NEO-FFI questions would be generated from a uniform distribution (also known as continuous or rectangular distribution), being one in which all values occur with equal frequency – that is, one in which all the possible answers are equally probable (Krysicki, Bartos, Dyczka, Królikowska, & Wasilewski, 2007). This assumption is made by a majority of researchers dealing with the computer simulation of random responding (Karabatsos, 2003).

Method

The Polish version of Costa and McCrae's NEO-FFI Personality Inventory (Zawadzki, Strelau, Szczepaniak, & Śliwińska, 1998) was used in the study; this questionnaire measures the intensity of five personality traits: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Each NEO-FFI scale is characterized by an acceptable accuracy of measurement (internal consistency). The values of Cronbach's α coefficient obtained in this study ranged from .59 to .82.

Full responses to all the 60 items of the NEO-FFI, which were qualified for analysis, were obtained from 943 of 996 participants who took part in the study. The participants were first-year undergraduate nursing students from nine Polish medical universities (Białystok, Bydgoszcz, Katowice, Cracow, Lublin, Łódź, Poznań, Warsaw, and Wrocław) that offer education at the undergraduate and graduate levels. A majority of the participants (80.1%) were between 19 and 21 years of age. A vast majority of the respondents (92.5%) were women. Data were collected while investigating the motives for taking up nursing studies (Kądalska & Fronczyk, 2006). The criterion data juxtaposed with these empirical data were generated from the uniform distribution simulating the responses of 1,000 individuals to each item of the NEO-FFI.

In order to verify the diagnostic capabilities of the indices, three logistic regression equations were computed, where the dependent variable was belonging to the category of either random or empirical data. In the first two equations, the two indices were treated as predictors separately; in the third equation, a combination of the indices was treated as a predictor. It was assumed that the value of the index, for which the probability of belonging to the category of random responders was .5 or higher, would classify a given person to that particular responding category. For each of the indices separately and for the combination of them both, the following were also computed: the percentage of correct classifications, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and Cohen's kappa agreement coefficient. Sensitivity is understood as the probability that a person displaying a particular responding style will be correctly classified (correct approvals). Specificity is the probability that a person responding in accordance with self-knowledge will not be diagnosed as exhibiting a particular style (correct rejections). NPV index is the probability that among the people diagnosed as responding in accordance with self-knowledge there are individuals who really respond in this way, whereas PPV index is the probability of random responding in a person diagnosed as a random responder

(Baer et al., 1997; Pinsoneault, 2002). Cohen's kappa coefficient compares the consistency of classification based on the analyzed index and the actual status as random or empirical data with the corresponding consistency in a situation of no relationship between these two classifications (i.e., with the participants randomly assigned to one of the categories) (Zieliński, 2008). On the basis of these calculations, it was determined which of the indices was more valid. The Receiver Operating Characteristic (ROC) curve was drawn for it. The ROC curve illustrates the relationship between the probability of false alarm and the probability of diagnosing random responding correctly for all possible cut-off values. It is constructed in the following way: the vertical axis represents the probability of diagnosing random responding correctly, whereas the horizontal axis represents the probability of false alarms. The classification power of the test is defined by the area under the ROC curve. Its maximum numerical value is 1, which means accurate classification in all possible cases.

All calculations were performed in the R computational environment using the basic set of packages. Only the values of Cronbach's alpha were computed using the psych package and the ROC curve was drawn using the code given by Ćwik and Mielniczuk (2009). The code generating random data after using the runif command is provided in Appendix 1. Appendix 2 contains the computation code for both random responding indices.

Results

First, the calculations connected with developing the sabotage index will be presented. Table 1 presents the results of five regression analyses for the halves of each scale. As can be seen, in the case of all scales the halves are significantly related to each other. In all cases, the value of the *t*-statistic, used for assessing the significance of the (β) regression coefficient, and the corresponding level of probability *p* confirm that this coefficient differs significantly from zero.

In a further step, residuals from the regressions described above were computed for each person. The sum of squared residuals gives the sabotage index. The parameters of the regression equation obtained in the sample of participants were used to determine the sabotage index for random data. Fixed individualized chance score was computed in accordance with the procedure described earlier. It required no estimations of any parameters for the whole sample.

The sabotage index and fixed individualized chance score are relatively weakly intercorrelated. In the group of random data, their correlation is not statistically significant ($r = .03$); in the case of data collected from participants, this correlation is significant ($p < .001$) but its value is not high ($r = .19$) and, with such a large sample, can hardly be regarded as proving any real relationship between the discussed variables.

Table 1
The Results of Five Regression Analyses for NEO-FFI Half-Scales

	Regression coefficient	Standard error	*t*	*p*
Dependent variable: Neuroticism – odd-number items				
Intercept	4.22	0.31	13.60	< .005
Neuroticism – even-number items	0.69	0.03	27.33	< .005
Dependent variable: Agreeableness – odd-number items				
Intercept	7.32	0.36	20.09	< .005
Agreeableness – even-number items	0.54	0.02	22.37	< .005
Dependent variable: Extraversion – odd-number items				
Intercept	7.14	0.37	19.49	< .005
Extraversion – even-number items	0.55	0.02	23.05	< .005
Dependent variable: Openness – odd-number items				
Intercept	9.98	0.45	22.25	< .005
Openess – even-number items	0.39	0.04	11.08	< .005
Dependent variable: Conscientiousness – odd-number items				
Intercept	5.88	0.39	15.20	< .005
Conscientiousness – even-number items	0.65	0.02	27.53	< .005

Next, (Table 2), the distributions of the two indices and the differences between their means were analyzed according to the type of data (actual or random).

The obtained distributions of both indices diverge considerably from the symmetrical distribution. The divergence is particularly marked in the case of actual data. The means of both indices differ considerably between the two groups of data.

Table 2
Distribution Properties of Random Responding Indices

	Data	Mean	Standard deviation	Skew-ness	Kurtosis	Wilcoxon *W*	*p*
Sabotage index	actual	45.56	37.09	2.44	9.79	180,379	< .005
	computer-simulated	97.08	56.89	1.03	1.28		
Fixed individualized chance score	actual	759.44	200.2	1.02	0.98	431,056	< .005
	computer-simulated	530.67	87.6	0.49	0.31		

The next step was to determine three logistic regression equations, where the predictors were as follows: the sabotage index, fixed individualized chance score, and both these indices combined. The parameters of those equations are collected in Tables 3-5.

Table 3
Parameters of Logistic Regression Classifying Random Responding on the Basis of the Sabotage Index

	Regression coefficient	Standard error	*z*-test	*p*	Odds ratio
Intercept	-1.70	0.1	-16.59	< .005	
Sabotage index	0.03	0.001	18.26	< .005	1.03

Table 4
Parameters of Logistic Regression Classifying Random Responding on the Basis of Fixed Individualized Chance Score

	Regression coefficient	Standard error	*z*-test	*p*	Odds ratio
Intercept	8.75	0.41	21.45	< .005	
Fixed individualized chance score	-0.01	0.001	-20.98	< .005	0.99

Table 5
Parameters of Logistic Regression Classifying Random Responding on the Basis of the Sabotage Index and Fixed Individualized Chance Score

	Regression coefficient	Standard error	*z*-test	*p*	Odds ratio
Intercept	7.08	0.45	15.83	< .005	
Fixed individualized chance score	-0.01	0.001	-18.87	< .005	0.99
Sabotage index	0.03	0.001	15.37	< .005	1.03

The obtained results show that the higher the value of the sabotage index and the lower the value of fixed individualized chance score, the more probable it is that a given person responded randomly. This is indicated by the positive parameter values for the sabotage index and negative values for fixed individualized chance score. Such a result was obtained with a high level of statistical significance when the two indices were considered separately (Tables 3 and 4) and in combination (Table 5). The values of odds ratios argue for similar conclusions. They are higher than 1 for the sabotage index and lower than 1 for the fixed individualized chance score. The divergences from 1 are not particularly high, but they are not decisive when it comes to the meaning of the discussed results. Both indices reach fairly high mean values and standard deviations, and therefore a value increase by a unit corresponds to a small increase in the probability of random responding.

The model fit values indicate that both the use of the sabotage index alone (*AIC* = 1,653.79) and the use of the fixed individualized chance score alone (*AIC* = 2,129.2) gives worse fit than the use of these two indices in combination (*AIC* = 1,279.57). The accuracy of classification based on each index separately and on the two indices in combination was assessed by computing a number of classification correctness measures. These are presented in Table 6, which shows that the classification based on the two indices used in combination yields the best results. This is indicated by the values of classification quality measures. Overall, 86% of all observations are classified correctly. Comparing the classifications based on the two indices used separately, it is difficult to decide, which of them is better. The fixed individualized chance score is characterized by a higher sensitivity, a slightly higher total number of correct classifications, and a better NPV index, but it gives a lower PPV index and a lower specificity than the sabotage index. The kappa coefficient indicates that the sabotage index is less accurate; the value is acceptable, though not particularly high (Zieliński, 2008). The value of kappa is somewhat higher in the case of the fixed individualized chance score, although it is still within the range of acceptable values. The combined use of the two indices gives a kappa value that is on the border between acceptable and high.

Table 6
The Accuracy Parameters of Random Responding Classification Based on Different Indices

	Sabotage index	Fixed individualized chance score	Sabotage index and fixed individualized chance score
Sensitivity	.69	.84	.87
Specificity	.79	.73	.85
PPV	.78	.77	.86
NPV	.71	.81	.86
% of correct classifications	.74	.78	.86
Cohen's kappa	.48	.57	.72

Summarizing, it seems that the best results in diagnosing random responding are obtained for the combined use of the sabotage index and the fixed individualized chance score. A ROC curve was drawn for the combination of the two indices (Figure).

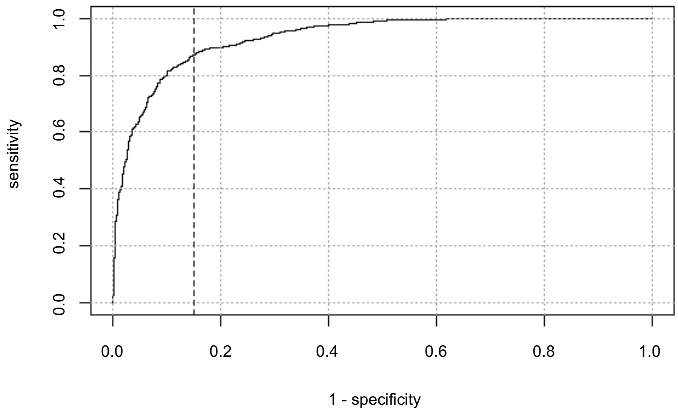

Figure shows that the indices used have a very high power of discrimination between random and actual data, which is indicated by the very strong bend of the curve. The area under the curve is .93, which means it is very large. The vertical dotted line in the figure indicates the specificity obtained in the study for the two indices in combination, namely .85. The chosen value of the logistic function (.5) above which a given pattern of responding is classified as random, seems to be moderately strict. The value of specificity is not particularly high compared to the value of sensitivity.

DISCUSSION

Both indices – the sabotage index and the fixed individualized chance score – have very skewed distributions, which is particularly visible in the case of data collected from actual participants. This can be easily understood, as it should be expected that most people respond in a consistent, nonrandom manner and so most values should point to responding in accordance with self-knowledge. It is puzzling that the sabotage index has a skewed distribution (although to a lesser degree) also for the computer-simulated data. In this case, symmetrical distribution should be expected, or a distribution skewed in the other direction, since higher values point to random responding. Perhaps such a result undermines the validity of the sabotage index a little. Still, the differences between the means of the two indices derived from actual data and simulated data reached a very high level of statistical significance.

The obtained results show that the exclusive use of the sabotage index according to Cattell et al. (1970) or of the fixed individualized chance score according to Haertzen and Ross (1978) offers moderate possibilities of identifying random responders. It is only the combined use of these two indices that gives much better results. This probably happens because the two indices measure somewhat different aspects of data randomness, as their weak intercorrelation suggests. The sabotage index is more sensitive to the lack of internal consistency of responses for each scale, whereas the fixed individualized chance score is related to the proportion of each response category, accounting for individual preferences for particular responses.

In other studies, it has been demonstrated that the combined use of several indices gives better results than the use of one index when it comes to the correct classification of responding styles or random responding. This tendency was found in the case of the MMPI (Archer & Elkins, 1999; Baer, Kroll, Rinaldo, & Ballenger, 1999). Hence, the result obtained in the presented study is not unusual. It may reflect the general tendency concerning the difficulties in distinguishing responding in accordance with self-knowledge from random responding.

The study yielded a fairly high proportion of correct classifications. An even higher proportion could be expected, but neither the actual data nor the simulated data were truly “pure.” This is because there may have been participants, who responded in a random way for some reasons, and among the computer-generated data there may have been cases of response configurations that, accidentally, happened to be combinations that could be the actual responses of actual participants. This partial ambiguity (even if it concerned only a minor part of the data) might have contributed to wrong classifications.

The study yielded a fairly high proportion of correct classifications. An even higher proportion could be expected, but neither the actual data nor the simulated data were truly “pure.” This is because there may have been participants, who responded in a random way for some reasons, and among the computer-generated data there may have been cases of response configurations that, accidentally, happened to be combinations that could be the actual responses of actual participants. This partial ambiguity (even if it concerned only a minor part of the data) might have contributed to wrong classifications.

As already mentioned in the introduction, the sabotage index would probably be the most accurate method of identifying random responders in the case of questionnaires with high internal consistency. Simulation research has shown that random responding diminishes the internal consistency of questionnaires (Fong, Ho, & Lam, 2010). Moreover, it may be supposed that for inventories consisting of a larger number of scales, this index offers better possibilities of distinguishing random responders from those responding in accordance with self-knowledge. The larger the number of scales, the lower the probability of many accidental divergences occurring among nonrandom responses and of those responses getting misclassified as random. Likewise, a larger number of scales and items allows better identification of random responding against responding in accordance with self-knowledge in the case of the fixed individualized chance score.

As suggested in the test manual (Zawadzki et al., 1998) for the Polish adaptation of the NEO-FFI questionnaire used in this study, the internal consistency of this questionnaire scales ranges from .68 to .82. Fairly similar values were obtained in the presented study. These values are not particularly high. This, in combination with a fairly low number of five scales, may have contributed to the moderately good result in diagnosing random responding.

The most important finding of this study is the demonstration that it is possible to distinguish between random responding and responding in accordance with self-knowledge quite effectively. What is more, the presented data make such diagnosing possible in psychological practice. Admittedly, this would involve laborious calculations, but using a spreadsheet will facilitate them considerably.

However, the practical application of these findings requires additional verification. The cut-off point values determined in various studies are very often different. This results from the specificity of samples and from the specificity of random data simulation. In order to apply the discussed indices in practice, it would be necessary to conduct additional cross-validation research or to use bootstrapping, which is based on a repeated selection of subsamples of participants and the verification of random responding indices. The averaged result of these multiple analyses will be the final result.

The presented study has certain limitations. First of all, the random data used were derived only from a uniform distribution. In reality, it is not certain if every alternative is chosen with the same probability. This is a simplification, since there may be a preference for one or another category of responses in purely random responding (van Ijzendoorn, 1984).

Another limitation may stem from the specificity of the participants in the presented study. As mentioned before, they were nursing students, predominantly women. It cannot be ruled out that young people and women have a stronger motivation to fill in questionnaires more reliably than the general population. This may have resulted in a lower percentage of carelessly completed questionnaires among the actual data, thus increasing the detectability of random responding.

Summing up, the very promising possibilities of identifying random or careless responding on the basis of the rather seldom used Cattell's sabotage index and the fixed individualized chance score should be highlighted. Further research – with other questionnaires and performed using more diversified samples obtained in various situations and with criterion data generated not only from uniform distribution – should contribute to the further verification of the diagnostic potential of both indices.

REFERENCES

Archer, R. P., & Elkins, D. E. (1999). Identification of random responding on the MMPI-A. *Journal of Personality Assessment, 73*, 407-421.

Baer, R. A., Ballenger, J., Berry, D. T. R., & Wetter, M. W. (1997). Detection of random responding on the MMPI-A. *Journal of Personality Assessment, 68*, 139-151.

Baer, R. A., Kroll, L. S., Rinaldo, J., & Ballenger, J. (1999). Detecting and discriminating between random responding and overreporting on the MMPI-A. *Journal of Personality Assessment, 72,* 308-320.

Beach, D. A. (1989). Identifying the random responder. *The Journal of Psychology, 123*, 101-103.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16PF).* Champaign, IL: IPAT.

Ćwik, J., & Mielniczuk, J. (2009). *Statystyczne systemy uczące się. Ćwiczenia w oparciu o pakiet R.* Warsaw: Oficyna Wydawnicza PW.

Evans, R. G., & Dinning, W. D. (1983). Response consistency among high F scale scorers on the MMPI. *Journal of Clinical Psychology, 39*, 246-248.

Fong, D. Y. T., Ho, S. Y., & Lam, T. H. (2010). Evaluation of internal reliability in the presence of inconsistent responses. *Health and Quality of Life Outcomes, 8*, 27.

Haertzen, C. A., & Ross, F. E. (1978). Using four chance profiles to estimate carelessness. *Psychological Reports, 41*, 1079-1087.

Irvine, M. J., & Gendreau, P. (1974). Detection of the fake 'good' and 'bad' response on the Sixteen Personality Factor Inventory in prisoners and college students. *Journal of Consulting and Clinical Psychology, 42*, 465-466.

Kądalska, E., & Fronczyk, K. (2006). Motywy wyboru studiów licencjackich na kierunku pielęgniarstwo w Polsce. *Pielęgniarstwo XXI wieku, 1-2*, 111-115.

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education, 16*, 277-298.

Krysicki, W., Bartos, J., Dyczka, W., Królikowska, K., & Wasilewski, M. (2007). *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach*. Wydawnictwo Naukowe PWN.

Nichols, D. S., & Greene, R. L. (1997). Dimensions of deception in personality assessment: The example of the MMPI-2. *Journal of Personality Assessment, 68*, 251-266.

Nichols, D., Greene, R., & Schmolck, P. (1989). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology, 45*, 239-250.

O'Dell, J. W. (1971). Method for detecting random answers on personality questionnaires. *Journal of Applied Psychology, 55*, 380-383.

Paluchowski, W. J. (1983). Źródła zakłóceń w kwestionariuszowym badaniu osobowości i ich kontrola. In W. J. Paluchowski (Ed.), *Z zagadnień diagnostyki osobowości* (pp. 249-271). Wrocław: Zakład Narodowy im. Ossolińskich.

Pinsoneault, T. B. (2005). Detecting random, partially random, and nonrandom Minnesota Multiphasic Personality Inventory-Adolescent protocols. *Psychological Assessment, 17*, 476-480.

Pinsoneault, T. B. (2002). A Variable Response Inconsistency scale and a True Response Inconsistency scale for the Millon Adolescent Clinical Inventory. *Psychological Assessment, 14*, 320-330.

Ross, F. E., & Haertzen, C. A. (1979). The use of chance profiles for detecting carelessness: The effect of determining the true response rate from items in scales. *Journal of Psychology, 101*, 27-35.

van Ijzendoorn, M. H. (1984). Answers without questions: A note on response style in questionnaires. *Perceptual and Motor Skills, 59*, 827-831.

Zawadzki, B. (2006). *Kwestionariusze osobowości: strategie i procedura konstruowania*. Warsaw: Scholar.

Zawadzki, B., Strelau, J., Szczepaniak, P., & Śliwinska, M. (1998). *Inwentarz Osobowości NEO-FFI Costy i McCrae. Adaptacja polska. Podręcznik.* Warsaw: Pracownia Testów Psychologicznych PTP.

Zieliński, A. (2008). Błąd klasyfikacji w badaniach epidemiologicznych. *Przegląd Epidemiologiczny, 62*, 461-470.

APPENDIX 1

PROGRAM R CODE GENERATING RANDOM NEO-FFI DATA

```
dane.losowe<-matrix(NA,1000, 60)
for (i in 1:60)
{
 dane.losowe[,i]<-runif(1000,0,1)
}
odpowiedzi.losowe<-
ifel-
se(dane.losowe<=1/5,0,ifelse(dane.losowe<=2/5,1,ifelse(dane.losowe<=3/5,2,ifelse(dane.
losowe<=4/5,3,4))))
```

APPENDIX 2

PROGRAM R CODE FOR COMPUTING THE SABOTAGE INDEX AND FIXED INDIVIDUALIZED CHANCE SCORE

```
lm.neu<-lm(neu_nparz~neu_parz)
lm.ekstr<-lm(ekstr_nparz~ekstr_parz)
lm.otw<-lm(otw_nparz~otw_parz)
lm.sum<-lm(sum_nparz~sum_parz)
lm.ugo<-lm(ugo_nparz~ugo_parz)
sabotowanie<-
lm.neu$residuals^2+lm.ekstr$residuals^2+lm.otw$residuals^2+lm.sum$residuals^2+lm.u
go$residuals^2
zlicz1<-rowSums(NEO_FFI== 1)/60
zlicz2<-rowSums(NEO_FFI== 2)/60
zlicz3<-rowSums(NEO_FFI== 3)/60
zlicz4<-rowSums(NEO_FFI== 4)/60
zlicz5<-rowSums(NEO_FFI== 5)/60
FIC_E=8*(zlicz2+2*zlicz3+3*zlicz4+4*zlicz5)+4*(4*zlicz1+3*zlicz2+2*zlicz3+zlicz4)
FIC_N=8*(zlicz2+2*zlicz3+3*zlicz4+4*zlicz5)+4*(4*zlicz1+3*zlicz2+2*zlicz3+zlicz4)
FIC_O=5*(zlicz2+2*zlicz3+3*zlicz4+4*zlicz5)+7*(4*zlicz1+3*zlicz2+2*zlicz3+zlicz4)
FIC_S=8*(zlicz2+2*zlicz3+3*zlicz4+4*zlicz5)+4*(4*zlicz1+3*zlicz2+2*zlicz3+zlicz4)
FIC_U=4*(zlicz2+2*zlicz3+3*zlicz4+4*zlicz5)+8*(4*zlicz1+3*zlicz2+2*zlicz3+zlicz4)
D=(neu-FIC_N)^2+(ekstr-FIC_E)^2+(otw-FIC_O)^2+(sum-FIC_S)^2+(ugo-FIC_U)^2
```