Modern Psychological Studies

Volume 22 | Number 2 Article 2

2017

The problem with online data collection: predicting invalid responding in undergraduate samples

Patricia Al-Salom
University of Windsor

Carlin J. Miller
University of Windsor

The Problem with Online Data Collection: Predicting Invalid Responding in Undergraduate Samples

Patricia Al-Salom and Carlin J. Miller[1]
University of Windsor

Abstract

The popularity of online research is increasing but the validity of the results obtained is not yet clear. The purpose of this study was to examine the factors that influence the validity of computerized data collection in an undergraduate sample. Participants were 99 university students randomly assigned to one of three data collection conditions: online survey platform, in-person computerized survey platform, and in-person pencil-and-paper survey. Results from statistical analyses suggest self-reported inattention symptoms, exposure to more stressors, and computerized platforms predict more invalid responding. In contrast, personality, self-reported impulsivity symptoms, and shorter completion times do not predict invalid responding. Overall, more than half of the participants failed at least one validity check and 11% failed three or more validity checks. Researchers, particularly those working with undergraduate samples, should consider implementing procedures to ensure the data collected are valid.

Keywords: online research, data validity, undergraduate students

The use of online data collection has risen in popularity over the past decade and reflects ongoing changes in the research process. Prior to the widespread use of questionnaire and survey data delivered by postal mail, participants either came into research labs or researchers traveled to their participants. By the 1970s, phone surveys became more popular whilst other studies continued to use postal service to transport data to and from participants, but the lack of anonymity was an issue in both cases (Alwin & Krosnick, 1991). With the ubiquity of the internet, a new avenue for collecting research became available. Today, internet-based research is commonplace and online surveys are considered cheaper, faster, and more convenient methods for accessing participants. Indeed, with online opportunities such as Mechanical Turk, researchers have access to samples that are vast, diverse, and motivated to respond to surveys.

Online data collection makes it easier to collect large-scale data very quickly and receive information from difficult-to-reach and traditionally underrepresented populations, such as Aboriginals or minorities (Andrews, Nonnecke, & Preece, 2003). Being able to contact far-flung participants with international collaborations or maintenance of longitudinal studies is also made easier with internet-based research (Dillman, 2007). By allowing individuals to participate online, participants may complete surveys in whatever setting they choose and thus they may be more likely to disclose information that they would otherwise be uncomfortable revealing (Bonini Campos et al., 2011). Thus, more accurate reporting rather than socially desirable reporting may result in contrast to what often happens when data is conducted in person (Aust, Diedenhofen, Ullrich, & Musch, 2013). This may be especially important in populations where individuals engage in high-risk behaviors, such as drug use or illegal activities (Barratt, Ferris, & Lenton, 2014). Other, more incidental advantages include reducing paper usage, postage costs, and the use of space for paper file storage (Fallaize et al., 2014). Based on the evidence of benefits through online data collection, it is clear why it is becoming more popular.

Although internet-based research clearly benefits researchers, the validity of the data collected is unclear. Online data collection is assumed to provide anonymity and therefore participants are more likely to respond candidly and genuinely, there is evidence to suggest this is not always the case (Aust et al., 2013; Hardre, Crowson, & Xie, 2012; Ihme et al., 2009; Oppenheimer et al., 2009; Ward & Pond, 2015). There are likely a number of factors that influence the quality of data provided. The physical disconnection from the researcher may increase the likelihood of careless responding (Hardre, Crowson, & Xie, 2012). The presence of a researcher in the room with the participant may also play a role in their performance, as evidenced by data from a study that randomly assigned participants to a room with a researcher present or a room with no researchers present (Burnham & Hare, 2007). Results from that study suggest that participants answer more carefully when in the same room as a researcher. There may also be personality or attitudinal differences that contribute to the validity of data. For example, Aust and colleagues (2013) observed a difference between those who described themselves as "serious" about answering the research survey and those who did not: self-described serious participants answered attitudinal and behavioral questions more consistently and predictably than non-serious participants. The time taken to complete items may also play a role in the validity of the data, because those who are rushing to complete the measures quickly may be more likely to respond carelessly (Ihme et al., 2009; Ward & Pond, 2015). There is also evidence to suggest that up to 10% of undergraduates participating in research studies as part of their coursework may apply suboptimal effort in responding to surveys (DeRight & Jorgensen, 2015). Thus, there are a number of factors that may play a role in the validity of data collected online.

It is critical that researchers are able to detect potentially invalid data in order to reduce noise within analyses. There are numerous methods for detecting these types of problems, including consistency checks, completion time monitoring, and instructional manipulation checks. With consistency checks, the consistency of responses across items is evaluated (Aust et al, 2013). For example, across measures or items, there may be multiple questions about test anxiety wherein it is implausible for an individual to report high levels of anxiety on one question and low levels on another item about test anxiety. This strategy is like to be more effective when the content of the items is heavily overlapping and there is little elapsed time between questions. Other studies exclude participants who have extremely short completion times (e.g., Ihme et al., 2009). This strategy is based on the assumption that those who finish very quickly are more likely to skim over instructions, not carefully consider their responses, and answer randomly to complete the survey as quickly as possible. Yet, it is difficult for researchers to determine what might be considered a "too short to be valid" completion time. A third strategy, the instructional manipulation check, embeds questions within the experimental material that ask participants to provide confirmation that they have read the questions within the study such as "please select "strongly agree"" for this answer (Oppenheimer et al, 2009). Regardless of the strategies employed, researchers must also use caution in removing participant data from a study as it reduces the power to detect effects and the winnowing of a dataset may influence results significantly, leading to Type I or Type II error. Researchers may also erroneously remove participants who represent diversity within the sample.

From our review of the literature on online data collection, it is not clear how often validity checks are employed in research currently. A meta-analytic review of online studies reported that out of 32 studies, only 6% reported the use of one or more measures of checking for the validity of data collected (Aust et al., 2013). Likewise, there are very few studies that have examined what factors predict invalid responding. The purpose of the present study was to evaluate the roles that personality, symptoms of inattention and impulsivity, exposure to hassles/stressors, and data collection method (online, in-person computerized data collection, or paper-and-pencil tasks with identical questions) play in the validity of data collected. In order to check validity of data, we used completion time monitoring and instructional manipulation checks. We hypothesized that those with lower Conscientiousness scores, higher Neuroticism scores, higher self-reported inattention, higher self-reported impulsivity, more stressful life events, shorter survey completion time, and those who completed surveys online would fail more validity checks.

Method

Participants

Participants were 99 undergraduate students (72.7% female) enrolled in one or more Psychology classes in an English-language Canadian university. In those who reported their ethnicity, 57% described themselves as Caucasian, 20% endorsed "other or mixed race," 8% Arab or of Arab descent, 5% Black/African-descent/Caribbean-origin, 5% Asian or of Asian descent, and 5% Hispanic. The sample was comprised mostly of 3rd (23%) and 4th (50%) year or beyond students. Of those reporting their major, 25% were Psychology majors, 53% reported other majors in the Arts, Humanities, and Social Sciences, 11% were Science majors, 8% were Business majors, and 3% were Human Kinetics majors. There were no exclusionary or inclusionary criteria for individuals to participate and our participants are largely representative of the department where the study took place.

Procedure

Participants were made aware of the study through the department's research pool and once they expressed interest in participating, they were randomly assigned to one of three data collection conditions. The online condition had 34 participants (26 females), the computerized condition had 34 participants (25 females), and the paper-and-pencil condition had 31 participants (20 females). Those in the computerized and paper-and-pencil conditions were scheduled for their informed consent and data collection in a lab with a researcher present during their entire participation. Each of those sessions had only participants assigned to the same condition (i.e., paper-and-pencil vs. computerized) in the room. Each participant was given a cubicle space to answer the items privately. Those in the online condition completed their informed consent and data collection entirely online, and had no in-person contact with the researchers. Those in the online and computerized data collection conditions completed their measures on a Fluid Surveys platform, which also calculated their time to complete the measures.

For all data collection conditions, the recruitment materials and the consent forms did not disclose that one of the central questions in this study was the influence of response format on data validity. All consent forms specifically noted that the surveys contained items to check if participants were reading all of the items. Following their participation, all participants received a letter of information form that explained the full purpose of the study, giving participants the opportunity to have their data removed from the study without penalty at that time. None of the participants requested to be withdrawn. As required by the ethics governing the research pool, participants received 0.5 bonus points in an eligible class of their choice for their 30 minutes of participation regardless of the validity of their data.

Each of the measures (listed below) had at least one validity check question randomly embedded, with a total of 7 validity checks in the study. These questions prompted participants to select a particular option (e.g., "please select 'strongly disagree' for this option". These were intended to check to see if individuals are reading the questions within the study and have been used in a variety of validity research experiments (Oppenheimer et al., 2009).

Measures

The measures described below were part of a larger battery. The measures not described are beyond the scope of the present study. Descriptive statistics (means and standard deviations) for all measures appear in Table 1.

Big Five Inventory (BFI). The BFI (John, Donahue, & Kentle, 1991) is a measure to designed to assess the five main personality characteristics in individuals, as described by Five Factor Theory of Personality (Costa & McCrae, 2003): Openness-to-New-Experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism. This self-report measure includes contains 45 items to which the participant responds on a five-point Likert scale ranging from "Strongly disagree" to "Strongly agree". Each scale is scored individually. The BFI has strong convergent validity, discriminant validity and test-retest reliability (Gosling, Rentfrow, & Swann, 2003).

ADHD-RS-IV with Adult Prompts (ADHD-RS). The ADHD-RS is a commonly used measure of ADHD symptoms for adults, namely inattention, hyperactivity, and impulsivity (Adler & Cohen, 2004). This self-report measure has 18 items to which the individual responds using a four-point Likert scale indicating the degree to which each symptom is apparent in their usual behavior (none, low, moderate, severe). These items match the DSM diagnostic criteria for ADHD. There are summary scores for Inattention items and Hyperactivity-Impulsivity items, as well as Total Symptoms. This measure has been shown to have substantial cross-cultural validity, internal consistency, convergent validity and test-retest reliability (Döpfner et al., 2006).

Inventory of College Students' Recent Life Experiences (ICSRLE). The ICSRLE is a self-report form about recent life events that may be described as stressors. It includes 49-items to which participants rate themselves on a four-point Likert scale, with responses of: (1) not at all part of my life, (2) only slightly part of my life, (3) distinctly part of my life, or (4) very much part of my life. The items in this measure relate to academic challenges, relationship issues, friendship problems and other life hassles and is used to assess exposure to stressful events in post-secondary students. The original norming sample data suggested adequate internal consistency (Chronbach's alpha = 0.89) and test-retest reliability (r = .825), with strong evidence for construct validity (Kohn et al., 1990).

Results

All analyses were performed using SPSS version 21.0. Visual examination of the data prior to analyses suggested the vast majority of participants failed no more than 3 of the validity checks and the groups beyond those who failed three items would be too small to analyze the group-level data. Based on our visual inspection of the data, we assumed three checks versus four checks and so on would not reflect meaningful differences in the groups. Thus, we formed three groups: "no fails", "one or two fails" and "three or more fails" for all of the analyses. Distribution of actual validity check failures is depicted in Figure 1. Data that were missing were not imputed (less than 2% missing data).

Contrary to our hypotheses, personality scores were unrelated to the number of validity checked failed. Using an omnibus MANOVA, groups could not be differentiated by number of items failed (Wilks' $\Lambda = 1.17$, $p = .32$). Thus, individual analyses of the factors (Conscientious, Agreeableness, Openness, Neuroticism, and Extraversion) were unwarranted. Likewise, time to complete the full survey did not differ significantly across the three groups ($F = 0.43$, $p = .65$).

Symptoms of inattention and hyperactivity-impulsivity produced mixed results. Inattention significantly differentiated the groups ($F = 5.49$, $p = .006$, $\eta^2_{partial} = .102$) but Hyperactivity-Impulsivity did not ($F = 1.98$, $p = .14$). On closer inspection, the group with no failures had significantly fewer inattention items endorsed at the moderate or severe level than the group with one or two failures ($p = .004$) but the group with three or more failures could not be differentiated ($p = .42$ - $.69$) from the other groups.

Exposure to hassles and stressors also significantly differentiated the groups. Those with no failures reported experiencing significantly fewer stressors in the last six months ($F = 3.51$, $p = .03$, $\eta^2_{partial} = .071$) than the other groups. Like the contrast analyses with inattention items, follow-up contrasts were not significant ($p = .43$ - $.95$).

Initial χ^2 analyses of response condition (online vs. computerized in lab vs. paper-and-pencil) suggested that the groups (O fails vs. 1-2 fails vs. 3+ fails) could not be differentiated based on response condition ($\chi^2 = 6.04$, $p = .20$). After examining the distribution of participants across the cells, we conducted *post hoc* analyses with new groups: those answering items on a computerized platform versus those answering on pencil and paper. The results from the *post hoc* analyses suggested that those answering questionnaires on a computer (whether in the lab or elsewhere) had significantly higher failure rates on the validity checks ($\chi^2 = 5.87$, $p = .05$, $\varphi = .24$).

Discussion

The goal of this study was to examine the factors that play a role in invalid responding by research participants. Results from the study suggest that having attention problems, experiencing an elevated level of life stressors and hassles, and responding to survey questions on a computer were associated with failing more validity check questions in an undergraduate sample. In contrast to our hypotheses, personality, self-reported hyperactivity-impulsivity, and the time spent completing the surveys was not associated with failing more validity checks.

Our results related to attention problems and exposure to hassles are not surprising. By definition, having difficulty paying attention, particularly to tasks that may be perceived as boring or when participants are not intrinsically motivated, should impact performance on a questionnaire. Our results suggest that the effect for attention problems is medium to large in size. There are numerous studies (e.g., Grane, Endestad, Pinto, & Solbakk, 2014; Ralph, Thomson, Cheyne, & Smilek, 2014) that suggest individuals, even those with subclinical attention problems, are more prone to errors.

Likewise, there are a number of studies that suggest that individuals experiencing distress are more likely to overlook details and make errors (e.g., Houston & Allt, 1997). It is not our intention to suggest that individuals who report attention problems or who have experienced stressful events should be excluded from research studies. Rather, we believe that because many individuals may have difficulty paying attention or may have elevated exposure to hassles, researchers may wish to include items to ensure invalid data is not included in research analyses.

The results from this study are in-line with a number of extant studies already published that suggest online research does not always result in valid data collection (Aust et al., 2013; Burnham & Hare, 2007; Hardre, Crowson, & Xie, 2012; Ihme et al., 2009; Oppenheimer et al., 2009; Ward & Pond, 2015). Indeed, like previous published work, our results suggest that more than 10% of our participants contributed data that is unlikely to be valid (DeRight & Jorgenson, 2015). Although online research may make recruiting participants and collecting data simpler, it may also be that certain populations provide data that is less valid. But, our work is not entirely in agreement with other studies. For example, both Ihme and colleagues (2009) and Ward and Pond (2015) reported that those with shorter completion times were less likely to contribute valid data. We did not find an effect for this in our results: participant completion times, whether in-person or using an online survey platform elsewhere, were highly similar across our sample. Examination of our data suggests that those who had more validation check failures completed the survey slightly faster (a difference of less than three minutes); thus, it may be that in a larger sample or with a more time-intensive survey, completion times may better differentiate those with invalid data from those with valid data.

Similar to the work by Burnham and Hare (2007), we used data collection conditions where the participant was in the room with a researcher and when there was no researcher present. Contrary to Burnham and Hare's results, we did not find an effect for researcher presence. Our results suggest that the computerized survey (whether on-line or in the lab) has a significant effect in the outcome of the data's validity. Notably, none of our participants who completed the measures on paper failed more than three of the validity checks, which was in contrast to those who accessed the survey on a computer.

Limitations

Although we are reporting significant results, there were several limitations in our study. Our sample size limited the available power to detect smaller effects and to detect differences between the sub-groups within our sample. This may have been particularly important in our ability to perform group contrasts in our analyses related to attention problems and exposure to hassles as some of the cell sizes were very small in those analyses. Future studies investigating invalid responses, particularly in undergraduate research pools, may benefit from larger sample sizes. Our survey was relatively brief in nature, which may have also limited our ability to test the effects of impulsivity, personality, and time spent on responding to survey items in detecting invalid responding. Similarly, with a longer survey, we would have also been able to use consistency checks to evaluate invalid responding. Our sampling procedure (i.e., using only those students enrolled in an undergraduate research pool within a Psychology department) resulted in a sample that was disproportionate in terms of number of females to males sampled. Likewise, because we used a university-based sample, our results may not generalize to other populations who may be participating in on-line research. We also did not record the amount of time participants took completing the surveys on paper due to the logistical issue of groups of participants completing surveys *en masse.* Lastly, no data were removed from our sample prior to analyses. Thus, it may be that the data from those participants with more errors may have biased our results toward or away from significance. Despite the limitations of our work, we believe the reasoning to include validity checks in online and computerized research platforms is solid.

Significance and Future Directions

Results from our study suggest that individuals who participate in online research may not contribute valid data, particularly if they are currently experiencing more stressful life events or if they have subclinical attention problems. As undergraduate students increasingly may fit into either of these two groups (Adlaf, Gliksman, Demers, & Newton-Taylor, 2001; Culpepper, 2011), research with this population may benefit from the inclusion of validity checks, using consistency checks or instructional manipulation checks, to ensure the data that is collected is accurate. Although online data collection is used more frequently now than ever before, results from our study suggest that it is critical to consider the quality of the data being collected.

References

Adlaf, E.M., Gliksman, L., Demers, A., & Newton-Taylor, B. (2001). The prevalence of elevated psychological distress among Canadian undergraduates: Findings from the 1998 Canadian campus survey. *Journal of American College Health, 50,* 67-72.

Adler, L., & Cohen, J. (2004). Diagnosis and evaluation of adults with attention-deficit/hyperactivity disorder. *The Psychiatric Clinics of North America*, *27*(2), 187–201. doi:10.1016/j.psc.2003.12.003

Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement: the influence of question and respondent attributes. *Sociological Methods & Research*, *20*(1), 139–181. doi: 10.1177/0049124191020001005

Andrews, D., Nonnecke, B., & Preece, J. (2003). Electronic survey methodology: A case study in reaching hard-to-involve Internet users. *International Journal of Human-Computer Interaction, 16* (2), 185-210.

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, *45*(2), 527–535. doi: 10.3758/s13428-012-0265-2

Barratt, M. J., Ferris, J. A., & Lenton, S. (2015). Hidden Populations, Online Purposive Sampling, and External Validity: Taking off the Blindfold. *Field Methods*, *27*(1), 3–21. http://doi.org/10.1177/1525822X14526838

Bonini Campos, J. A. D., Zucoloto, M. L., Sampaio Bonafé, F. S., Jordani, P. C., & Maroco, J. (2011). Reliability and validity of self-reported burnout in college students: A cross randomized comparison of paper-and-pencil vs. online administration. Computers in Human Behavior, 27(5), 1875-1883. doi: http://dx.doi.org/10.1016/j.chb.2011.04.011

Burnham, T. C., & Hare, B. (2007). Engineering Human Cooperation: Does Involuntary Neural Activation Increase Public Goods Contributions? *Human Nature*, *18*(2), 88–108. doi: 10.1007/s12110-007-9012-2

Costa, R.R. & McCrae, P.T. (2003). Personality in adulthood : A Five Factor Theory Perspective. New York : Guilford Press.

Culpepper, L. (2011). Recognizing and diagnosing ADHD in college students. *Journal of Clinical Psychiatry, 72, e33*. doi: 10.4088/JCP.11009tx2c.

DeRight, J. & Jorgenen, R.S. (2015). I just want my research credit: Frequency of suboptimal effort in a non-clinical healthy undergraduate sample. *Clinical Neuropsychologist, 29,* 101-117. doi: 10.1080/13854046.2014.989267.

Dillman, D. A. (2007). *Mail and internet surveys: the tailored design method* (2nd ed., 2007 update with new internet, visual, and mixed-mode guide). Hoboken, N.J: Wiley.

Döpfner, M., Steinhausen, H.-C., Coghill, D., Dalsgaard, S., Poole, L., Ralston, S. J., … ADORE Study Group. (2006). Cross-cultural reliability and validity of ADHD assessed by the ADHD Rating Scale in a pan-European study. *European Child & Adolescent Psychiatry*, *15 Suppl 1*, I46–55. doi: 10.1007/s00787-006-1007-8

Fallaize, R., Forster, H., Macready, A. L., Walsh, M. C., Mathers, J. C., Brennan, L., … Lovegrove, J. A. (2014). Online Dietary Intake Estimation: Reproducibility and Validity of the Food4Me Food Frequency Questionnaire Against a 4-Day Weighed Food Record. *Journal of Medical Internet Research*, *16*(8), e190. doi: 10.2196/jmir.3355

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in Personality, 37, 504-528.

Grane, V.A., Endestad, T., Pinto, A.F., & Solbakk, A.K. (2014). Attentional control and subjective executive function in treatment-naive adults with Attention Deficit Hyperactivity Disorder. *PLoS One, 29*, e115227. doi: 10.1371/journal.pone.0115227.

Hardre, P. L., Crowson, H. M., & Xie, K. (2012). Examining Contexts-of-Use for Web-Based and Paper-Based Questionnaires. *Educational and Psychological Measurement*, *72*(6), 1015–1038. doi: 10.1177/0013164412451977

Houston, D.M. & Sophie K. Allt, S.K. (1997). Psychological distress and error making among junior house officers. *British Journal of Health Psychology, 2,* 141-151.

Ihme, J. M., Lemke, F., Lieder, K., Martin, F., Müller, J. C., & Schmidt, S. (2009). Comparison of ability tests administered online and in the laboratory. *Behavior Research Methods*, *41*(4), 1183–1189. doi: 10.3758/BRM.41.4.1183

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory--Versions 4a and 54. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

Kohn, P. M., Lafreniere, K., & Gurevich, M. (1990). The Inventory of College Students' Recent Life Experiences: a decontaminated hassles scale for a special population. *Journal of Behavioral Medicine*, *13*(6), 619–630.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. doi: 10.1016/j.jesp.2009.03.009

Ralph, B.C., Thomson, D.R., Cheyne, J.A., & Smilek, D. (2014). Media multitasking and failures of attention in everyday life. *Psychological Research, 78*, 661-699. doi: 10.1007/s00426-013-0523-7.

Ward, M. K., & Pond, S. B. (2015). Using virtual presence and survey instructions to minimize careless responding on Internet-based surveys. *Computers in Human Behavior*, *48*, 554–568. doi: 10.1016/j.chb.2015.01.070

Appendix

Table 1

Descriptives (means and standard deviations) for each measure by group

	Total Sample N = 99	0 fails n = 43	1-2 fails n = 45	3+ fails n = 11
BFI				
Conscientiousness	26.35 (3.43)	27.00 (2.91)	25.57 (3.73)	27.25 (3.73)
Agreeableness	25.02 (3.28)	24.53 (2.86)	25.13 (3.91)	27.00 (2.14)
Openness	30.40 (3.24)	30.28 (3.60)	30.49 (2.78)	30.50 (3.96)
Neuroticism	25.27 (4.16)	25.58 (3.85)	25.44 (4.49)	22.63 (3.20)
Extraversion	25.41 (3.61)	25.28 (3.55)	25.38 (3.70)	26.25 (3.85)
Inattention **(ADHD-RS)**	2.62 (2.82)	1.79 (1.61)	3.33 (2.56)	3.00 (2.45)
Impulsivity-Hyperactivity **(ADHD-RS)**	1.83 (2.10)	1.35 (1.65)	2.20 (2.34)	2.20 (2.66)
Life stressors **(ICSRLE)**	92.51 (22.54)	85.88 (17.63)	98.28 (24.81)	95.89 (25.66)
Time to complete in hrs.*	.34 (.21)	.33 (.15)	.36 (.24)	.29 (.29)
Response format				
% online	34	28	36	55
% computer	34	33	33	45
%paper	32	39	31	0

** Did not include those completing the measures in the paper-and-pencil condition. Completion time also reflected the larger battery, much of which was beyond this particular study.*

Figure 1.
Percentage of sample failing number of validity items.