Response latencies are alive and well for identifying fakers on a self-report personality inventory: A reconsideration of van Hooft and Born (2012)

Ronald R. Holden · Christine E. Lambert

Abstract Van Hooft and Born (*Journal of Applied Psychology* 97:301–316, 2012) presented data challenging both the correctness of a congruence model of faking on personality test items and the relative merit (i.e., effect size) of response latencies for identifying fakers. We suggest that their analysis of response times was suboptimal, and that it followed neither from a congruence model of faking nor from published protocols on appropriately filtering the noise in personality test item answering times. Using new data and following recommended analytic procedures, we confirmed the relative utility of response times for identifying personality test fakers, and our obtained results, again, reinforce a congruence model of faking.

Keywords Response latencies · Faking · Congruence · Response favorability

Can the response times associated with answering self-report personality items identify individuals who are faking? Over the decades, various perspectives regarding response times and dissimulation have emerged. Some have argued that faking is a complex process that, relative to honest answering, requires extra cognitive processing and editing. In support of this perspective, McDaniel and Timm (1990) found that, when instructed to answer dishonestly, respondents took more time to answer items than when they were instructed to answer honestly, particularly for items pertaining to socially undesirable behaviors (e.g., drug use). Furthermore, in advocating for a response editing approach

R. R. Holden · C. E. Lambert
Queen's University, Kingston, Ontario, Canada

R. R. Holden (✉)
Department of Psychology, Queen's University, Kingston,
Ontario K7L 3N6, Canada
e-mail: holdenr@queensu.ca

to answering, Holtgraves (2004) had respondents answer personality items in contexts varying in evaluation demands, and found that instructions inducing socially desirable responding resulted in longer response times than did other instructional sets.

A contrary view, whereby faking is associated with faster responding, also has empirical support. For example, Holden, Fekken, and Jackson (1985) reported that shorter latencies were associated with answering items that were more saturated with socially desirable responding. Hsu, Santelli, and Hsu (1989) found shorter item response times for fakers than for respondents answering under standard instructions. Similarly, George (1990) indicated that, as compared to honest responding, faking was associated with shorter response latencies.

Given the empirical evidence that faking can seemingly result in either slower or faster responding, a more complex model becomes necessary. Nowakowska's (1970) schematic diagram for responding to questionnaire items proposes multiple pathways that consider both the social desirability of item stimuli and a cognitively demanding intellectual evaluation associated with response disclosure. As such, socially desirable responding (e.g., faking) can be more primitive, easy, and straightforward (i.e., fast), or can be more controlling of the truth and, thus, slower. Holden, Kroner, Fekken, and Popham (1992) have also articulated a more complex model of faking that is derived from schema theory. Their model indicates that whether fakers respond faster or slower depends on the congruence between the generated response and the faking schema. In particular, dissimulators faking in the direction of positivity will provide favorable responses (i.e., congruent answers) faster than they provide unfavorable (i.e., noncongruent) responses. Complementarily, dissimulators faking in the direction of negativity will provide unfavorable (i.e., congruent answers) faster than they will provide favorable (i.e., noncongruent) responses. Thus, noncongruent responding is

slower than schema-congruent answering. What is important is not the positivity or negativity of the item (i.e., positively or negatively keyed items) per se, but rather, the positivity or negativity of the answer and how this response relates to a schema for faking. This distinction between keying and responding is detailed in the Appendix. One might question why fakers provide answers that are noncongruent with their faking schema; however, fakers do produce such noncongruent responses (e.g., Holden, 1995). It may be that faking can be a nuanced, sophisticated process rather than merely simple, naïve answering and that, consequently, fakers will seek to avoid presenting an extremely obvious and detectable dissimulation pattern.

Empirical support for the congruence model has been found in studies of experimentally induced faking with university students (Brunetti, Schlottmann, Scott, & Hollrah, 1998; Esser & Schneider, 1998; Holden et al., 1992), incarcerated offenders (Holden & Kroner, 1992), and unemployed persons who are actively seeking employment (Holden, 1998). For evaluating this congruence model, Holden and colleagues (Holden, Fekken, & Cotton, 1991; Holden & Hibbs, 1995; Paulhus & Holden, 2010) supplied specific analytic procedures (including a figural representation) that detailed appropriate methods for addressing the noise endemic in raw response latencies for individual test items (Fazio, 1990). First, in addition to addressing outlier item response times, latencies are standardized within a respondent in order to control for irrelevant person sources of variance, such as reading speed and motor speed. Next, latencies are standardized within an item in order to control for irrelevant item sources of variance, such as item length and vocabulary level. Then, latencies are aggregated within a respondent, bearing in mind the nature of the respondent-generated answer (e.g., positive vs. negative response).

Recently, in purportedly testing the Holden et al. (1992) congruence model of faking, van Hooft and Born (2012) indicated null results for the model and reported (p. 312) that "No support was found for Holden et al.'s interactive model of faking." Furthermore, in their (alternative) approach to analyzing response latencies for the identification of fakers, van Hooft and Born reported mixed results across dimensions of personality, with only an overall small effect size emerging. Their response latencies did have significant added value, but enhanced the correct identification of fakers and honest respondents just marginally, from a 77.5% to a 79.8% hit rate (vs. a chance value of 50%).

Closer inspection of the van Hooft and Born (2012) article, however, reveals multiple potential issues with their analyses. Item response latencies were not standardized within participants. As such, the variances in van Hooft and Born's

latencies still retained the presence of irrelevant person factors such as reading speed, verbal ability, and motor speed.[1] Also, item response latencies were not standardized within items. As such, variances in van Hooft and Born's latencies still retained the presence of irrelevant item factors such as item length, item ambiguity, and item extremity. Although van Hooft and Born indicated that their item latencies were aggregated, their method of aggregation did not align with either the model or the procedures outlined by Holden et al. (1992). Whereas van Hooft and Born aggregated item latencies by separately calculating average response latencies for positively and negatively *keyed* items, this is not in accord with Holden et al. (1992) model, which distinguishes between positive (or favorable) and negative (or unfavorable) *answers*. The congruence model regards the favorability or unfavorability of the generated *response* as critical. As such, although van Hooft and Born appeared to fail to replicate Holden et al.'s (1992) model, it was a mistaken null result for that model.

Is the Holden et al. (1992) model a stable, legitimate phenomenon that reflects nontrivial effect sizes? Although replicable findings (Holden, 1998; Holden & Hibbs, 1995; Holden & Kroner, 1992) from multiple laboratories (Brunetti et al., 1998; Esser & Schneider, 1998) have indicated so, van Hooft and Born's (2012) recent, nonoptimally analyzed data in the *Journal of Applied Psychology* suggest otherwise. Here, we acquired and analyzed new data to either confirm or refute the congruence model. In doing so, we followed the detailed analytic procedures advocated by Holden and associates (Holden et al., 1991; Holden & Hibbs, 1995; Paulhus & Holden, 2010).

Method

Materials

The stimulus materials consisted of the NEO-FFI (Costa & McCrae, 1992), a 60-item self-report measure that assesses the five-factor personality model (Neuroticism, Openness, Agreeableness, Conscientiousness, and Extraversion; 12 items per scale). Costa and McCrae have reported NEO-FFI scale coefficient alpha reliabilities above .73 and validities, based on correlations with spousal report, above .33.

---

[1] van Hooft and Born (2012) used a within-subjects design in their faking manipulation, and although this could be regarded as controlling for individual respondent differences, the use of such a design in faking studies is problematic because of the presence of differential carryover effects, whereby the effects of faking are more extreme when it follows rather than precedes honest responding (Holden, 1997).

Participants and procedure

The participants were 293 undergraduates (227 women, 66 men) recruited either through an introductory psychology course subject pool or by posting flyers on campus. Individuals were compensated with either course credit or $15. The mean age of the sample was 18.82 years (SD =1.04, range 17 to 24).

The NEO-FFI items were computer-administered, and participants were asked to answer the research materials as if they were being screened for military induction, under one of three randomly assigned instructional conditions: (1) standard instructions, (2) fake responses to maximize their chances of being inducted (i.e., fake good), or (3) fake answers to minimize their chances of being inducted (i.e., fake bad). All respondents were warned of the presence of validity checks to detect faking, were asked to do their best to avoid being detected, and were given an incentive to do so: For each 25 participants, a $50 prize was awarded to the participant who was farthest from activating the validity checks.

Preparation of response latencies

The NEO-FFI Neuroticism scale was reverse-scored, so that higher scores indicated higher levels of emotional stability, and thus, higher scores on all of the NEO-FFI scored dimensions (emotional stability, openness, extraversion, agreeableness, and conscientiousness) now represented favorable (as opposed to unfavorable) responding. Subsequently, in accord with the recommended procedures (Holden, 1998; Holden & Kroner, 1992; Holden et al., 1992), raw response latencies were adjusted to reduce the effects of statistical outliers and were standardized twice, once to control for confounding person variables (e.g., reading speed and gender) and a second time to control for confounding item variables (e.g., length, complexity). This multistep standardization procedure was completed as follows: First, to reduce the effect of statistical outliers, response latencies were Winsorized, such that values of less than 0.5 s or greater than 40 s were set to 0.5 or 40 s, respectively.[2] Second, response times were standardized across items within each participant in order to control for differences between individual participants. Third, response times were standardized across participants within each item in order to control for differences between items. Importantly, the mean and standard deviation used for this standardization were computed using only the response time data from participants in the honest condition. Finally, items were re-Winsorized, such that latencies of less than −3.00 or greater

than 3.00 were set to −3.00 or 3.00, respectively. This double-standardization procedure produces response latencies that are calibrated both with respect to the person and with respect to the test item and are free of the confounding main effect influences of individual persons, individual items, and statistical outliers.

Once item latencies were standardized and Winsorized, item response times were then aggregated for each participant in order to produce means for keyed responses (agreeing or strongly agreeing with a favorable/positive item) and also means for nonkeyed responses (disagreeing or strongly disagreeing with a favorable/positive item). For each participant, these two mean adjusted response latency scores were the units of analysis. Item latencies for "neutral" responses were excluded from the analyses.

Results

As a manipulation check for the faking instructions, a multivariate analysis of variance on NEO-FFI scale scores indicated significant differences among the instructional groups, Wilks's λ = .220, $F(10, 572)$ = 64.80, $p < .001$, η_p^2 = .53. In particular, more favorable responses were present for the group maximizing their chances of military induction than for the standard instruction group, Wilks's λ = .583, $F(5, 190)$ = 27.20, $p < .001$, η_p^2 = .42, and fewer favorable responses were given by the group minimizing their chances of military induction than by the standard instruction group, Wilks's λ = .365, $F(5, 189)$ = 65.78, $p < .001$, η_p^2 = .64. Thus, substantial dissimulation with more than a large effect size had been successfully induced.

Analysis of response latencies based on Holden et al. (1992)

Mean adjusted response latencies as a function of instructional condition are presented in Table 1. Using Endorsement (favorable vs. unfavorable) as a within-subjects factor and Faking Instructional Group as a between-subjects factor, a significant faking by endorsement interaction (Fig. 1), $F(2, 289)$ = 39.13, $p < .001$, η_p^2 = .22, approximating a large effect size (i.e., .25; Murphy & Myors, 2004), confirmed the Holden et al. (1992) congruence phenomenon in terms of both statistical significance and meaningful magnitude.[3] Using Tukey's HSD, post-hoc comparisons among faking groups within each endorsement level (favorable or unfavorable) indicated that favorable responses for the fake bad group were significantly slower than those for the fake good or the standard instruction group, which did not differ. For unfavorable responses, the

---

[2] Appropriate upper and lower limits for Winsoring are somewhat arbitrary and can depend on factors such as item length and complexity and a respondent's reading ability and motor dexterity. We reanalyzed the data using a 10 s rather than a 40 s upper limit and obtained virtually identical results.

---

[3] In the analysis of interaction effects, main effects of endorsement and faking condition were still included and statistically modeled.

Table 1 Mean (and *SD*) adjusted response latencies by faking condition

Type of Response	Instructional Group			$F(2, 289)$
	Standard (*n* =98)	Fake Good (*n* =97)	Fake Bad (*n* =97)	
Positive/favorable responses	−0.09 (0.12)	−0.10 (0.10)	0.16 (0.33)	48.04[*]
Negative/unfavorable responses	0.03 (0.32)	0.20 (0.49)	−0.10 (0.13)	17.98[*]

One participant in the fake good group produced no unfavorable responses and, consequently, was excluded from the analyses. [*] $p < .001$.

fake good group was significantly slower in answering than was the standard group, which was significantly slower than the fake bad group. These significant differences again supported and confirmed the Holden et al. (1992) congruence model of faking.

Using the two mean adjusted response latency scores, a discriminant function analysis (assuming equal prior probabilities) evaluated the ability of these latencies to correctly classify individuals into faking groups. Discrimination was significant, Wilks's $\lambda = .709$, $\chi^2(4, N =292) = 99.17$, $p < .001$; the overall correct classification hit rate was 60.27% (as compared to a chance rate of 33.33%); and the group correct classification hit rates were 63% (62 of 98), 56% (54 of 97), and 62% (60 of 97) for the standard, fake good, and fake bad groups, respectively.

Reanalysis of the present data using the procedure of van Hooft and Born (2012)

Response latencies were also analyzed using van Hooft and Born's (2012) procedure. That is, rather than Holden's (Holden et al., 1991; Holden & Hibbs, 1995; Paulhus & Holden, 2010) procedure of double standardization and subsequently aggregating separately for positive and negative *responses* (regardless of item keying), no double standardization was applied and aggregation was undertaken separately for positively and negatively *keyed* items (irrespective of response).

Using Item Keying (positively vs. negatively keyed) as a within-subjects factor and Faking Group as a between-subjects factor, a significant faking by keying interaction, $F(2, 291) = 10.46$, $p < .001$, $\eta_p^2 = .06$, represented a medium effect size. With Tukey's HSD, post-hoc comparisons among the faking groups within each item keying indicated that, whereas there were no differences among the faking groups in answering times for the negatively keyed items, the group instructed to fake bad was significantly slower to answer positively keyed items than were the other two groups, which did not differ significantly. Using the two mean response latency scores based on item keying (van Hooft & Born, 2012), a discriminant function analysis (assuming equal prior probabilities) was significant, Wilks's $\lambda = .896$, $\chi^2(4, N =294)$

$= 31.85$, $p < .001$. The overall correct classification hit rate was 45.57% (as compared to a chance rate of 33.33%), and the correct classification hit rates were 27% (26 of 98), 55% (54 of 98), and 55% (54 of 98) for the standard, fake good, and fake bad groups, respectively.

Discussion

Fazio (1990) has highlighted that, for response latencies, "the 'signal-to-noise' ratio can be far less than is necessary to detect the presence of a true effect statistically" (p. 77). Nevertheless, he also indicated that, when appropriately analyzed and interpreted, response times can have informational value. We concur with that perspective: Response latencies are influenced by a variety of factors, random and systematic, irrelevant and relevant, that produce a raw form that is potentially crude and confounded. Isolating out reliable information from this conglomerate can be a challenging process. On the basis of earlier cognitive models of test item processing (Holden et al., 1991; Rogers, 1974a, b, 1978), Holden et al. (1992) detailed a procedure for deriving latency information that would be relevant for detecting fakers. The success of that technique has been replicated in a variety of studies and laboratories. Here, we again confirmed that procedure and, now, also extended its application to an additional personality inventory (the NEO-FFI) and to five-point response ratings (i.e., *strongly disagree* to *strongly agree*), from previously studied items that had been answered true or false (e.g., Holden & Hibbs, 1995). The similarity between our graphical summary in Fig. 1 and the figures presented by Holden et al. (1992, their Figs. 1, 2, and 3) and by Esser and Schneider (1998, their Fig. 4) is particularly noteworthy and attests to a remarkable consistency of the congruence phenomenon.

The magnitude of the added value of using response latencies analyzed by means of the Holden et al. (1992) procedure is noteworthy. Our observed effect size (i.e., $\eta_p^2 = .22$) was very close to large, and the change, in accurately identifying fakers and honest respondents, from a chance level of 33.33% to 60.27% (i.e., a difference of 26.94%) constitutes an improvement of 80.83% (i.e., 26.94/33.33) from baseline. This added value contrasts markedly with van Hooft and Born's
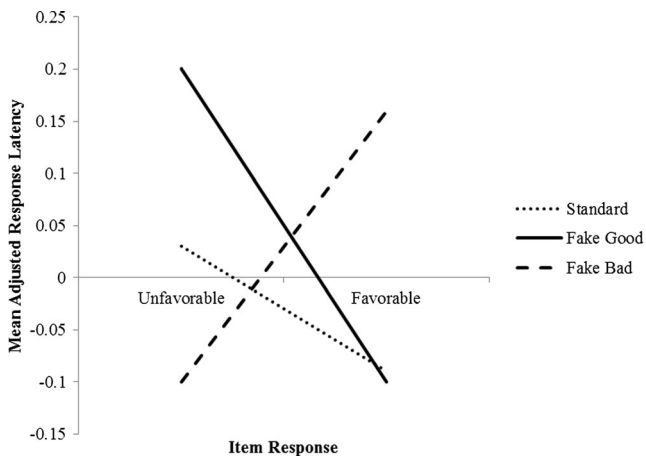
Fig. 1 Mean adjusted response latencies, by response favorability and faking condition

(2012) data, in which suboptimal analyses produced a significant, but small, effect size (i.e., Cohen's $d = 0.23$) and a change in the accurate classification of faking and honest respondents from 77.5% to 79.8% (i.e., 2.3%), which represented an improvement of only 2.96% (i.e., 2.3/77.5) from baseline. Although perhaps not directly comparable statistically, our presently obtained 80.83% improvement dwarfs the significant, but minor, 2.3% enhancement of van Hooft and Born, again highlighting the nonoptimal nature of their analyses. When the present data were analyzed using van Hooft and Born's method of analysis, relative to a chance hit rate of 33%, the overall classification hit rate of 46% was substantially smaller than the 60% hit rate achieved through the Holden et al. (1992) analytic procedure. Therefore, response latencies, when appropriately analyzed, can have substantial added value.

The discrepancies between the nature and magnitude of the results found when the present response latency data were analyzed using the methods of Holden et al. (1992) and van Hooft and Born (2012) highlight that the two analytic applications may address different processes. In presenting new data that replicated both the effect and the size of the effect for the Holden et al. (1992) method of analysis, we do not dispute the findings of van Hooft and Born (2012), but rather argue that their analyses neither test the Holden et al. (1992) model of personality test item response dissimulation nor fully optimize the use of response times for the identification of fakers on self-report personality inventories. In their 2012 study, van Hooft and Born generally found that (p. 308) "faking involves a faster response process for both positively and negatively *keyed* items" (our emphasis). This finding aligns with that of others (e.g., George, 1990; Hsu et al., 1989; Nowakowska, 1970) and indicates that, in general, presenting positively may be an automatized component of self-presentation. Paulhus (1993), in describing the process of responding to personality items, asserted that an association between positivity of

responding and automaticity exists and that it may be attributable to a tendency toward positive self-presentation taught from childhood.

Although such a finding manifested in item keying can represent an interesting phenomenon, the effect and its analyses do not adequately address Holden et al.'s (1992) model of personality test item faking, which emphasizes the importance of the match between a respondent's answer and his or her schema for faking. In that model, responses that are congruent with a faking goal will be faster than responses that do not match that faking goal. This approach, derived from schema theory, views personality scale item response latencies as self-schema indicators that demonstrate an inverted-U effect, whereby respondents scoring high or low on a trait respond more quickly than do moderate scorers on that trait (Akrami, Hedlund, & Ekehammar, 2007). Relative to moderate scorers on a personality dimension, higher scorers endorse a trait-relevant item more quickly, and lower scorers reject the trait-relevant item more quickly (Holden et al., 1991). When adapted to a faking schema, relative to honest respondents, this means that individuals faking good will endorse positivity faster and negativity slower, whereas persons faking bad will endorse positivity slower and negativity faster (Holden et al., 1992).

Potential limitations to our research exist. First, although experimental instructions to fake produced large effect sizes for group scale scores, there may have been individual participants who did not fake, even with the offered incentive. A second possible issue arises in our treatment of emotional stability, extraversion, openness, agreeableness, and conscientiousness as positive characteristics. This constellation of five traits collectively is seen by some as a general factor of personality (Rushton & Irwing, 2008) that can be interpreted as a general dimension of desirability (Backstrom, Bjorklund, & Larsson, 2009). Our choice of a military induction scenario represented a context in which respondents could reasonably be expected to fake either positively, for induction, or negatively, to avoid induction. Nevertheless, the positivity or negativity of individual traits for faking goals may vary with the specific goal and could be tailored to be more specific than is summarized by a general factor of personality.

It might be asked whether potential strategies could enable faking and yet elude detection by response times. In principle, from our findings, we speculate that these strategies do exist. One such strategy would be to answer all items in the direction of the faking goal. However, although this approach might evade response latency detection, it would result in scale scores that would be readily detectable by an experienced test interpreter or by a standard validity/lie scale. A second strategy would be for a respondent to avoid slowing down when providing an item response that is contrary to the faking goal. Whether, in practice, such a strategy can be implemented by respondents remains a direction for future research.

In summary, the merit of using response latencies aggregated according to a congruence model of faking is alive and well. However, when response times are not aggregated appropriately, suboptimal analyses can lead to attenuated results that underestimate the utility of response latencies and that can mistakenly challenge a valid model of response dissimulation.

Appendix: Distinguishing item keying from item answering in calculating mean item response latencies

Consider four items for a scale of sociability, a desirable characteristic, answered by two respondents (Robert and Elizabeth) who either endorse or reject each of the items. Items 1 and 4 are positively keyed for sociability; Items 2 and 3 are negatively keyed.

	Robert	Elizabeth
1. I am friendly.	Endorse	Endorse
2. I am unfriendly.	Reject	Reject
3. I am hostile.	Reject	Endorse
4. I am not hostile.	Endorse	Reject

The van Hooft and Born (2012) analysis calculates a mean for positively keyed items using the response times for Items 1 and 4 (regardless of the respondent's answers) and a mean for negatively keyed items using Items 2 and 3 (again regardless of the respondent's answers). Holden et al.'s (1992) model, however, distinguishes between positive and negative responses. As such, Robert provides answers indicative of the desirable characteristic of sociability for each item, and therefore his mean response time for favorable answers is based on all four items. His mean latency for unfavorable responses cannot be calculated, because he provided no unfavorable answers. Elizabeth gives responses indicative of the desirable characteristic of sociability only for Items 1 and 2, and therefore her mean response time for favorable items is based on two items (Items 1 and 2). Her mean latency for unfavorable responses is based on the latencies for Items 3 and 4. The Holden et al. (1992) analysis does require additional response time preparatory filtering (for a detailed example, see Paulhus & Holden, 2010) prior to calculation of the means.

References

Akrami, N., Hedlund, L.-E., & Ekehammar, B. (2007). Personality scale response latencies as self-schema indicators: The inverted-U effect revisited. *Personality and Individual Differences, 43,* 611–618.

Backstrom, M., Bjorklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality, 43,* 335–344.

Brunetti, D. G., Schlottmann, R. S., Scott, A. B., & Hollrah, J. L. (1998). Instructed faking and MMPI-2 response latencies: The potential for assessing response validity. *Journal of Clinical Psychology, 54,* 143–153.

Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R professional manual: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factory Inventory (NEO-FFI).* Odessa, FL: Psychological Assessment Resources.

Esser, C., & Schneider, J. F. (1998). Differentielle reaktionslatenzzeiten beim bearbeiten von persönlichkeitsfragebogen als möglicher indikator für verfälschungstendenzen [Differential response latencies as a possible indicator for detecting faking on personality test items]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 19,* 246–257.

Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clarke (Eds.), *Research methods in personality and social psychology* (pp. 74–97). Newbury Park, CA: Sage.

George, M. S. (1990). *The use of response latency to study accuracy and consistency in a computerized lifestyle assessment.* Unpublished doctoral dissertation, University of Toronto, Toronto, Ontario, Canada.

Holden, R. R. (1995). Response latency detection of fakers on personnel tests. *Canadian Journal of Behavioural Science, 27,* 343–355.

Holden, R.R. (1997, June). *Studying faking on personality tests: Statistical and research design issues.* Paper presented at the Canadian Psychological Association Annual Convention, Toronto, Canada.

Holden, R. R. (1998). Detecting fakers on a personnel test: Response latencies versus a standard validity scale. *Journal of Social Behavior and Personality, 13,* 387–398.

Holden, R. R., Fekken, G. C., & Cotton, D. H. G. (1991). Assessing psychopathology using structured test item response latencies. *Psychological Assessment, 3,* 111–118.

Holden, R. R., Fekken, G. C., & Jackson, D. N. (1985). Structured personality test items and validity. *Journal of Research in Personality, 19,* 386–394.

Holden, R. R., & Hibbs, N. (1995). Incremental validity of response latencies for detecting fakers on a personality test. *Journal of Research in Personality, 29,* 362–372.

Holden, R. R., & Kroner, D. G. (1992). Relative efficacy of differential response latencies for detecting faking on a self-report measure of psychopathology. *Psychological Assessment, 4,* 170–173.

Holden, R. R., Kroner, D. G., Fekken, G. C., & Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology, 63,* 272–279.

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin, 30,* 161–172.

Hsu, L. M., Santelli, J., & Hsu, J. R. (1989). Faking detection validity and incremental validity of response latencies to MMPI subtle and obvious items. *Journal of Personality Assessment, 53,* 278–295.

McDaniel, M. A., & Timm, H. W. (1990). *Lying takes time: Predicting deception in biodata using response latency.* Paper presented at the 98th Annual Convention of the American Psychological Association, Boston.

Murphy, K. R., & Myors, B. (2004). *Statistical power analysis* (2nd ed.). Mahwah, NJ: Erlbaum.

Nowakowska, M. (1970). A model of answering to a questionnaire item. *Acta Psychologica, 34,* 420–439.

Paulhus, D. L. (1993). Bypassing the will: The automatization of affirmations. In D. M. Wegner & J. W. Pennebaker (Eds.), *Handbook of mental control* (pp. 573–587). Englewood Cliffs, NJ: Prentice Hall.

Paulhus, D. L., & Holden, R. R. (2010). Measuring self-enhancement: From self-report to concrete behavior. In C. R. Agnew, D. E. Carlston, W. G. Graziano, & J. R. Kelly (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 221–240). New York, NY: Oxford University Press.

Rogers, T. B. (1974a). An analysis of the central stages underlying responding to personality items: The self-referent decisions and response selection. *Journal of Research in Personality, 8,* 128–138.

Rogers, T. B. (1974b). An analysis of the stages underlying the process of responding to personality items. *Acta Psychologica, 38,* 205–213.

Rogers, T. B. (1978). Experimental evidence for the similarity of personality and attitude items responding. *Acta Psychologica, 42,* 21–28.

Rushton, J. P., & Irwing, P. (2008). A general factor of personality (GFP) from two meta-analyses of the Big Five: Digman (1997) and Mount, Barrick, Scullen, and Rounds (2005). *Personality and Individual Differences, 45,* 679–683.

van Hooft, E. A. J., & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology, 97,* 301–316.