

Possible Effects of Social Science Measuring and Measurements on Persons Measured and Unmeasured: Reactivity

Merton S. Krause
Evanston, Illinois

Believing one or persons like one are being, were, or will be measured on some social or psychological dimension and by whom may influence how one participates in being measured or otherwise lives one's life in ways that affect one's such measurements and their generalizability. Believing that one knows one's own or certain others such measurements also may. Therefore the social sciences need to detect, avoid, and overcome these problems of *reactivity* to their measuring and measurements if they are to obtain valid measurements and generalizations from these. This requires the cooperation of the measured and so an understanding of the social psychology of measuring on such dimensions, of how persons participate in being measured on such dimensions and react to being informed about their own or some others measurements on these dimensions. These are matters that physical science measurement theory has no reason to be concerned with, is not, and so cannot properly provide a model for measurement of persons on social science dimensions.

Keywords: openness; impression management; consequences of measuring; consequences of the measurements

Psychological traits (T) are widely estimated from rather brief tests of a very special kind: persons' choices among prescribed alternative responses to each of several items. A key purpose of T is to predict some aspect of how *individual persons* will live their life as a basis for deciding what options to offer these persons, such as a job, scholarship, parole, etc. or for deciding how to intervene on how they do live their life, such as by retraining, tutoring, monitoring...them. Item sets are assembled for such tests (see, e.g., Cano, Vosk, Pendrill, & Stenner, 2016; De Ayala, 2009; Finkelstein, 2005, for thoughtful introductions to how, and Krause, 2017, for a critique).

The usual Psychometric approach to this predicting relies on correlations in samples of many persons' test scores with their criterion scores (see, e.g., Irvine, 2010; Strauss & Smith, 2009), rather than on examinations of the whole distribution of individuals' test and criterion score associations. This allows there to be (a) predictive validity at the overall sample correlation level, as indicated by a statistically significant correlation, conjointly with (b) extensive unpredictability at the individual test and criterion score association level (Krause, 2018; Lee, 1993). The logic of statistical significance testing allows increasingly smaller correlations to

be significant as sample size becomes larger, so the bigger the sample the smaller the correlation that can be statistically significant.

The possibility of some invalidly high scorers' (e.g., test cheaters or test taking sophisticates) scores compensating for invalidly low scorers' (e.g., distracted or overly anxious test takers) scores makes poor individual case prediction possible in conjunction with a statistically significant sample correlation. Only the detection of both these sorts of individual case score invalidity can reveal the extent of this problem. The attractiveness of the mathematical sophistication of Linear Model statistics has blinded many to the importance of the individual case, especially for Psychology (see Danziger, 1985; Krause, 2018).

Mass ability or disability screening for educational, psychotherapeutic, correctional, etc. intervention assignment or for policy formulation can sometimes be reasonably cost-effective despite such individual level invalidity, but the individuals measured and those who are to intervene on them (e.g., teachers, psychotherapists, parole officers, etc.) deserve valid information on each individual case rather than only statistically significant sample correlations. To the extent that $r < 1$ it cannot provide the necessary individual case guidance (Krause, Lutz & Böhnke, 2011). What may serve the customers for mass testing results well enough for their purposes need not serve the tested well enough for *their* purposes, and how the tested see their purposes served by some test may influence *how they participate* in the testing.¹

The Validity of a Trait Measurement

Several (m) items are generally used to measure persons on a trait (T) because (a) each item represents only a portion of the T and/or (b) each is affected by different influences that cancel each other out, respectively. So T are widely presumed in Psychometrics to be manifest in persons' m test-item responses as well as in how the tested persons subsequently live their lives, as manifest in the strength of the correlation between these two.

Such Psychometric predictions can be and at the time of testing can seem importantly consequential to the persons tested. Insofar as they do

¹ Producing comparable difficulty level items for each difficulty level for mass testing depends on averages and so is properly intended to inform some institutions policies regarding aggregations of persons rather than to validly measure individual persons. Having multiple versions of an item, each with a different "incidental" component for the same "radical" component, but all at the same *average* difficulty level, is used to obstruct the test takers' familiarity with one "incidental's" inflection of a "radical" from serving as familiarity with any other "incidental's" inflection of this "radical" (Irvine, 2010). This should impede coaching to favorably bias the scores of those coached until coaches learn about the alternative "incidentals". It also entails, however, that each individual person's test score is somewhat biased by the set of "incidentals" encountered and differently so on a retesting involving a different set of "incidentals".

seem so, the measureds' trying to obtain what then seem to them the most personally favorable T scores is to be expected. This will be constrained by each such person's resources for obtaining a favorable score and by their ethical or other qualms about how to do so. Thus, psychological measurement is not performed by measurers on measureds (or as physical measurement theorists put it "measurands") that are utterly unconcerned about and unreactive to what and how measurements are produced on them, which means persons are radically unlike the physical sciences' objects or events measured.

Some psychological measurement theorists (e.g., Mari, 2005 & 2013; Maul, Irribarra & Wilson, 2016) try (unsuccessfully: see, e.g., Finkelstein, 2005) to reconcile social science "soft system" (Checkland, 2000) measurement (which is "soft" because of persons reactivity to being measured) with "hard system" physical measurand measurement by assuming that "the objectivity [of a Measurement System: MS] implies that the MS is able to discriminate the measurand from the various [extraneous] influence quantities so that the acquisition component of the MS is sensitive only to the measurand..." (Wilson, 2013, p. 3773). To assume this is to fail to take account of how craftily *reactive* persons can be to being measured on social science dimensions, which shall be argued and documented below. This *reactivity* can affect each test, interview, or overt observing of persons and each retesting, re-interviewing, or re-observing of that person somewhat differently. To the extent that psychological measuring or measurements causally influence the measured in terms of the dimensions measured on, the validity of these measurements and of any generalizations of these to other persons will be biased (see e.g. Lee, 1993; Wetzell, Böhnke & Brown, 2016, on extremity, acquiescent, and social desirability biases; and Hewitt et al., 2003 on perfectionist biases). Therefore, unlike "hard system" measurement, psychological measurement requires the active cooperation of the measureds with the measurers, so the logic of "hard system" measurement (see, e.g., Luce & Narens, 2008; Mari, Carbone, Giordani, & Petri, 2017) is not properly applicable to social science measurement.²

Predicting from Test Scores

Predicting on the basis of test scores has become a rather big business (see, e.g., Carter, Daniels & Zickar, 2013; Gibby & Zickar, 2008; Hoffmann, 1964). It avoids having to rely on persons likely

² They may implicitly have allowed for *reactivity*, for "the ontological problem about the time-persistence of SuM [i.e., measurands]" in their footnote 4 and for measurand and measurement "system evolution" in their footnotes 7, 8, and 11. Although they do not explicitly deal with "soft system" measuring, they may have made some openings for doing so.

faulty and perhaps biased reflection on their own history or on inept or biased observers' descriptions of these persons for making predictions about these or other relevantly similar persons. A great variety of events are important for social science to be able to predict, such as high school and college graduation, behavior contrary to the public interest, becoming or remaining a productive employee, successful parent, compatible cohabiter...(see, e.g., Harano, Peck & McBride, 1975; Sackett, Borneman & Connelly, 2008).

Subjecting persons to Psychometric tests is obviously less costly and intrusive than meticulously gathering biographical data, but how persons respond to being measured biographically (Habermas, 2007) versus being measured by testing apparently remains unexplored. Differences in how persons respond to knowing the measurements obtained in these two ways apparently also remains unexplored, as also does the measured's suspicion or awareness of disclosure to some of the unmeasured (such as co-racials, co-religionists, co-professionals). These are kinds of *reactivity* to measuring and measurements unique to persons and so a serious measurement validity problem for the social sciences, a problem the physical sciences do not have.³

³ Thus, for the social sciences the process of measuring does not consist only of the measurers' actions on the measurand, as Rossi & Crenna (2016) implicitly require in their explication of the notion of a *measurement system*: "...for the measurement procedure to be really applicable – as it actually is – we should require that it ends with a transduction where the indication may be directly expressed by numbers."

Instead, for the social sciences the measured system, the human measurand, takes an active role in the *measurement system* for producing numbers, the final "transduction", descriptive of her or him self that suit her or his purposes rather than exclusively, if at all, the measurers' purposes.

Measuring persons on psychological dimensions properly involves a "soft" rather than "hard" *measurement system*, one that takes sufficient account of the measured as actively participating individual stakeholders with their individual stakes and ways of enhancing these. As Finkelstein (2005) puts it: "Systems that include human actors present, as far as their observation is concerned, a problem of self awareness. By this is meant, that if the system is observed, and the fact of the observation and its results are known, the human actors tend to alter their behaviour. The measurement [i.e., *measuring*] thus significantly distorts the observation." Shishkin (2016): "...measurements of non-physical quantities...have nothing to do with physical science metrological methodology." Maul, Mari, Irribarra & Wilson (2018) recognize that this disjuncture exists but suppose that eventually it can be overcome. They do not, however, take the measured's and un-measured's *reactivity* to the measuring or measurements at all into account. In the meantime the social sciences' still predominant preoccupation with the statistics of aggregations of persons (Krause, 2018) has abetted its measurement theories taking "hard" rather than "soft" *measurement system* form because person aggregations per se are not reactive only some of their constituent persons are reactive.

Measuring how Persons Live their Lives

A measurement is the instantiation of a gradation on some descriptive dimension, so measuring a person on several psychological dimensions locates the person at some point in a hyperspace defined by a set of mostly ordinally (Krause, 2012, 2013; Michell, 2008, 2010) gradated descriptive dimensions. Since every description of how a person is living their life at some particular time represents a conjunction of gradations across some set of dimensions, it necessarily depends upon measurements in terms of these dimensions' gradations. Such describing and so the measuring underlying it are quite ordinary everyday activities of persons.

Psychological measuring must seem to a measured person either (a) intentionally overt, (b) ineffectively covert, or (c) not done (i.e., effectively covert). The results of the measuring may be believed by a measured person to be revealed (d) only aggregately or (e) individually to (f) only the individual measured themselves, (g) also to the measurers' clients, (h) only to the measurers' clients, or (i) publicly, which may include members of the measureds' reference groups (Hyman & Singer, 1968; Van Praag, 2011), such as their relatives, friends, clients, employers. Each measured person may believe that any one of these 3 (a-c) times 2 (d-e) times 4 (f-i) possibilities is the case, and which of these a measured believes when being measured may influence her or his participation in the measuring and so may influence what measurements will result insofar as the person cares.

If one *cares* about what oneself or others believe about one, then believing that some of one's T have been, are being, or will be measured may induce one to participate in the measuring differently than if one does not care. Because there obviously are sometimes good reasons to care about who believes what about one and on what evidence, this is something about the measuring situation that the social sciences should take seriously. Caring can be about (j) having been (e.g., Goffman, 1978; Teghtsoonian, 1989), being (e.g., Fiske, 1971, pp. 203-227), or going to be (e.g., Sakaquaptewa & Thompson, 2003) measured on what T; (k) what the measurements were, are, or will be; and (l) about who knows or will know about these measurements. For persons who do care there is the issue of whether they (m) *change* how they live, including how they take tests, in order to obtain what they themselves, or certain other persons apparently would, take to be favorable measurements. Thus, social science measuring inherently involves these further influences that makes it different than physical science measuring with its measurands that do not know or care about being measured, what the measurements are, or who knows them nor that can try to change themselves because of these measurements.

Only overt social science measuring has so far been our concern here, so some consideration of how persons may participate in their overt measuring is in order now. After dealing with this and some issues arising from it, we shall turn to the nature of and issues in covert measuring.

Overt Measuring

Overtly psychologically measured persons obviously have an active role in their measuring that can involve trying to favorably influence the measurements obtained on them. This can be by lying, cheating, faking, cramming, imbibing stimulants or relaxants, seeking guidance on how to favorably participate in the measuring through learning about the interviewer, the test, the proctoring, the scoring of answers, how to use some test items or interview questions as cues to how others should be dealt with. Measurers, therefore, properly must take the measured's reacting to being measured (e.g., Miller & Barrett, 2008; Schwarz, 1999; Schwarz et al., 2008; Tourangeau & Bradburn, 2010) and the nature of the social situation in which the measuring is done (e.g., Bolino et al., 2008; Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001) into account in order to try to ensure that the obtained measurements are valid.

For one example, educational testing must deal with the possibilities of (a) students' bribing or coercing to see a test ahead of time, getting special coaching for it (e.g., Cohen, 2006; Cornett & Knight, 2009), querying previous test takers for advance information on it, sneaking relevant unlearned information into the testing situation or copying from other test takers (e.g., Cizek, 1999); and of (b) some teachers' teaching for the specific test they expect their students to be given (e.g., Meier & Wood, 2004; Monfils, Firestone, Hicks, Martinez, Schorr, & Camilli, 2004; Popham, 2001). For another example, faking (e.g., Walsh, 1990) on personality tests can be equally pragmatic (e.g., Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; DeMaio, 1984; Griffith & Peterson, 2006) and has been attempted to be countered by the measurers' including seemingly covert "social desirability" items (e.g., Ellis, West, Ryan, & DeShon, 2002; Nicholson & Hogan, 1990; Webb, Campbell, Schwartz, Sechrest, & Grove, 1981; Zickar & Gibby, 2006) or adapting Item Response Theory (e.g., Zickar & Drasgow, 1996) to detect it.

Some of the measured, however, may want valid measurements made to inform themselves about what these measurements are so that they can rely on the measurements to guide themselves in subsequently living their lives, instead of using the occasion of their being measured as an opportunity to manage only what others believe about them. For example, one's valid achievement test results can guide one as to whether and where further learning is needed by one, as to how educated or capable a person one can reasonably consider oneself, as to what future options one may

and may not have. Analogous interests exist for mental health status, personality trait, and employability measurements. Thus, in any sample of persons measured on some psychological dimension there will likely be some who seek only an apparently favorable measurement and some who seek only a valid measurement: two sorts of persons whose measurements properly are incomparable, making the average score for a sample that includes both sorts of limited or no practical use.

What the measured can learn from their being measured depends upon the nature of the measuring. Overt measuring may be opaque (O) or it may be transparent (~O) as to what *dimensions* are being measured on, whether intentionally made so or not as to each particular dimension the measured are being measured on (see, e.g., Nichols & Edlund, 2015). This may be as to what information (*o*) is obtained or inferred from the measurements on these dimensions, as to the means (*p*) by which this information is derived from the measuring on these dimensions, and as to what uses (*q*) this information will or may be put by whom. (These three matters differ from the 13, (a) – (m), already discussed above). The various configurations of these 16 may be felt by each person measured as (r) uncomfortable or not, (s) personally dangerous or not, (t) personally useful or not, and perhaps responded to accordingly, which likely will be different for different persons and so may variously confound the measurements obtained. Thus, social science test and interview data properly require analysis in all these regards for each person measured.

Each of these possibilities for each measured dimension has its own implications for both the measureds (see Bornstein, 2011, although he deals with the statistics of aggregated data rather than with individuals data) and the measurers participation in the measuring and for the validity of the individual measurements obtained. This validity depends on what the measured want to believe about themselves and to have others believe about them, on what the measured believe to be the nature of the measuring process, and on their preference for and skill at dissembling (e.g., Zickar & Robie, 1999; Ziegler, MacCann & Roberts, 2011) in the specific form of measuring applied. As yet the social sciences know far too little about all these matters, but individuals' psychological measurements are nevertheless too routinely unjustifiably treated as valid (e.g., Cizek, Rosenberg & Koons, 2008; Meyer et al., 2001), as if there were not a complex of psychological processes involved in being measured (as Bornstein, 2011, details and as the scientific normative definition of each dimension *should* detail: Krause, 2012).

Therefore, sophisticated measureds have to master the complex craft (Krause, submitted) of *being measured* in order to obtain valid measurements for their own purposes or to manage for their own purposes the measurements on them that others obtain for their purposes. This too poorly understood craft of *being measured* confronts the (in part because

of it) still too unappreciatedly complex psychological craft of *measuring* T in order for valid measurements to be obtained, both of which crafts influence what measurements are obtained on each *individual* person measured. This is what ultimately makes social science “soft system” measurement radically different than physical science “hard system” measurement. In other words, social science measuring often is a *competitive* inter-personal psychological craft somewhat driven by other objectives than simply the valid measuring of each individual measured. Social science ethics and individual-measurement validity, however, require that the measuring always be a *cooperative* inter-personal craft, the cooperation ideally based upon a confluence of both parties’ interests in obtaining valid measurements. This is a complication utterly foreign to physical measurement, which for this reason too cannot be a proper model for psychological measurement, since only properly sophisticated sentient measureds *can* properly cooperate with their measurers.

Effects of Publishing Results of Social Science Measuring

Learning (truly or falsely) from some publication of measurement results that persons whom one believes oneself to be like have on average certain disadvantages (e.g., that “Type A personalities” are prone to heart attack, “jocks” to later rheumatism or brain damage) or that persons whom one believes oneself to be unlike have on average certain advantages over persons like oneself (e.g., that extraverted activities are likely to reduce loneliness or that persons who carry concealed weapons are likely to feel less anxious than those who don’t) may induce one to try to somewhat change how one lives one’s life. Thus, learning about measurements made on others may evoke T changes in oneself that *countervail or conditionalize the generalizability of these measurements*. For example, because of what they have learned from published or somehow reported about studies, “Type As” may then engage in stress reduction practices that make them less prone to having heart attacks, “jocks” who then avoid excessive stress on their joints may become less likely to become rheumatic, introverts who then make themselves socialize more or make their aloneness spiritually or creatively fulfilling may avoid feeling lonely, the fearful who then arm themselves or realize that being armed merely trades one risk for other risks may moderate their fearfulness, etc. Thus, generalizations made from measurements on some sample of persons to a population they are some sample of may be falsified due to individuals’ responses in this population (including among the measured themselves) to publication of the results of the measuring. So far as I can tell this possibility remains mostly unexplored by the social sciences (except for election polls: see Marsh, 1984; Morwitz & Pluzinski, 1996; Moy & Rinke, 2012; Rothschild & Malhotra, 2014; Perse, 2001)

On these grounds too social science measuring is radically different than the traditional model of measuring on physical dimensions (see, e.g., Chang & Cartwright, 2008; Luce & Narens, 2008; Mari, 2005 & 2013), here in two different ways. First, this traditional model in effect assumes that the *measuring* directly affects either nothing about whatever population is studied (as in, e.g., Astronomy, Marine Geology, Meteorology, History) or affects only the particular individual cases it studies (as in, e.g., Mechanics, Quantum Physics, Chemistry, Botany, Biology of infra-humans). Second, it in effect assumes that the *measurements* obtained do not themselves affect the measured or the corresponding measurements of any of the unmeasured of that same population (see Bracha & Burkle, 2006; Burkle, 1996). Even Epidemiology, Clinical Medicine, Human Pharmacology can sometimes have both these *reactive* properties (see, e.g., Crocco, Villasis-Keever & Jadad, 2002; Kinsman, 2012). So it is unjustifiable for the social sciences to simply proceed as if their measuring or measurements obtained on any given dimension do not affect the persons overtly measured or those informed (truly or falsely) of the measuring in terms of, respectively, the measurements obtained on the measured or, by implication, those that may be obtained were some of the unmeasured who take themselves to be in the same relevant population as the measured to be similarly measured. This further distinguishes physical science from social science measurement and also calls for some consideration of covert psychological measuring and secrecy about the measurements obtained.

Covert Measuring and Secret Measurements

Some persons may prefer that they not be measured on certain T (e.g., intelligence, psychopathy, introversion, wealth...) because being measured would or might be disadvantageous by revealing things about them they prefer to be private or to not themselves know. Some may prefer that they seem not needing to be measured on certain T (e.g., PTSD, dishonesty, distractibility, patriotism...) because this favorably distinguishes them from those who apparently need to be measured on these T. Some may prefer that they not seem unmeasured on certain T (e.g., ambitiousness, logicity, stress tolerance, political party preference...) when others are measured on these, because this might make them seem unworthy of measurement on these T. Such preferences may influence which persons actually get measured and so what population is represented by the data obtained.

All this makes covert measuring of T a tempting option, which carefully crafted multi-factorial item sets and response-set detection items make feasible so long as the measured remain naïve about such devices. (Which cannot be long enough in an internet connected world of the measured.)

For those measured who are not naïve about such devices the detection or suspicion of covert measuring endangers their cooperation with their measurers that, as argued above, valid psychological measurement requires.

Thus, the mere fact as well as the nature of measuring persons on social science dimensions may be influential on the lives of the persons measured and, because they learn about the measuring or measurements, even on the lives of some of the unmeasured. This possible influence deserves itself to be measured although such measuring of the effects of measuring or not measuring may then have its own effects, etc. Could not all this be avoided by covert measuring where feasible? Not ethically in open democratic societies for public enterprises such as the social sciences insofar as the covert measuring itself or the measurements obtained might be harmful for some persons. Unless the measurements were clearly essential for the public good, as some epidemiological ones are if they do not evoke dangerous panic (as in Kinsman, 2012). Informal everyday psychological measurement provides a useful perspective on all this.

Informal Everyday Psychological Measurement

In contrast with such ethical constraints on Scientific Human Psychology (SHP), each one of us quite regularly and informally, covertly as well as sometimes overtly, measures others' T, and may disseminate information produced by such measuring. For example, such remarks as "I can see without her telling me that she's delighted at having aced that test."; "Just to see them together is enough to recognize their incompatibility."; "He is much less religious than he claims to be." etc. are commonplace. Persons' *informal* covert and overt psychological measuring are more loosely ethically constrained than is social science measuring and occurs in everyone's everyday life, with most everyone informally measuring themselves and others and also being informally measured by others on all sorts of dimensions, although most of this goes unnoticed, unremarked about, and unreflected upon (see Luft, 1961 on this).

Such informal psychological measuring in everyday life obviously differs in some respects from formal SHP measuring, in part because so much of it is or seems essential for each of us in living one's life. The measuring done under the auspices of the social sciences, however, primarily is concerned instead with the induction/grounding and the testing of generalizations about how persons live their lives and why so. These scientific measurements need to *validly* indicate gradations on the dimensions measured on and *usefully* indicate or predict matters that require detection or prediction (see, e.g., Teller, 2008).

The social need for certain valid and useful measurements makes some covert measuring attractive in everyday life and for the social sciences. However, insofar as ethical considerations require avoidance of covert social science measuring, overt measuring must be relied upon, which to be valid must involve the measureds' cooperative participation in the measuring. This means that it is necessary for these sciences to find ways of and dimensions for overt measuring on that evoke the cooperation of the measured, which may well require something of a cultural sea change because controlling what others believe about one is vitally important for many of us. Such control is a crucial (but unequally distributed) social skill, that of quotidian impression management, of opportunistically being opaque or dissembling.

The Measureds' Opacity and Dissembling

Often pitted against the craft of measurement, informal and formal, on social science dimensions is the craft of impression management (see, e.g., Bolino, Kacmar, Turnley, & Gilstrap, 2008; Cohen, 2006; Cozby, 1973; Dilchert, Ones, Viswesvaran, & Deller, 2006; Edwards, 1990; Ellis et al., 2002; Glaser & Strauss, 1964; Goffman, 1978; Griffith & Peterson, 2006; Hewitt et al., 2003; Krumpal, 2013; Leary & Kowalski, 1990; Paulhus, 2001; Pennington, Heim, Levy, & Larkin, 2016; Spencer-Oatey, 2007; Wetzel, Böhnke & Brown, 2016) that prevents the role complementarity between the measurer and the measured required for valid overt measuring (see, e.g., Broen & Wirt, 1958; Chan, 2014; *Human Performance*, 2011; Ziegler, MacCann & Roberts, 2011). Serious, competent, honest, and scientifically disinterested participation in being measured requires cooperation with and transparency to one's measurers with regard to this dimension. This may require much preparatory dialogue between measurers and measured (e.g., when interviewing is to produce the data: Sudman & Bradburn, 1982, pp. 29-118). Rational (as distinct from naïve) measureds will naturally require thorough transparency (i.e., overt openness) by the measurers about what dimensions are being measured on, what information is obtained or inferred from the measuring, the means by which information is derived from the measuring process, and to what uses the measurement information obtained will be put by whom (see, e.g., Posner, 1981; Smith, Dinev & Xu, 2011).

It may be difficult for a person to learn (or relearn) how to be utterly transparent to some others about what one is like, about how one does and means to live one's life. Each of us likely manages to be transparent/open sometimes to certain others about certain matters, but few if any of us are open to everyone about everything and so we sometimes practice the craft of impression management. Effective impression management requires

that one validly measure one's own T as well as some others' T as accurately as is feasible in order to try to choose correctly whom to be how open to when about what. In other words, to most strategically and effectively manage the measurements others formally or informally take on one, one ideally would preemptively take measurements on oneself and on these others, a very demanding preoccupation that always lacks certainty about its sufficient accomplishment. So we necessarily compromise on how hard to work at this with whom and when, but being subjected to overt social science measurement should certainly challenge those who are rational to decide how hard to work at impression management when being so measured. This further distinguishes social science "soft system" from physical "hard system" measurement.

Because unconditional or ill considered openness can have complicating consequences, such as endangering or disadvantaging oneself or others, openness is something to sometimes rationally replace with opacity or dissembling so that for each of us these must naturally vary over time and context (e.g., Hui & Triandis, 1985) and can have various objectives (e.g., Broen & Wirt, 1958). This limits others' and so the social sciences' ability to validly measure one on dimensions one means to and can effectively dissemble or be opaque on, which then can distort generalizing from measured to unmeasured persons. Thus, social science knowledge claims about the distribution of persons on any dimension will be biased by the (unlikely random) opacity and dissembling of at least some of the *measured*. So their measurements are not likely accurately generalizable to whatever population of persons of which they are some sample (Krause, 2016; Krause, Lutz & Böhnke, 2011). This makes it necessary to consider *measurers'* counteractive dissembling.

Measureds and Measurers Bilateral Dissembling

Living one's life in such a way that particular others perceive one to be living it in one way while one actually is living it in some other way constitutes a successful performance of the craft of dissembling. Mastering this craft of quotidian acting requires that the measured be clear about what impression to make on whom, when, and how to covertly do this. To once be seen as a dissembler on any dimension by a measurer is to risk having subsequent measurements on at least this dimension seen by this measurer and his or her colleagues as also likely invalid and measurements on other dimensions as at least suspect of being so.

Measureds' dissembling and opacity are meant to evade measurer attempts to detect them. So in order to be valid social science measuring must either (a) counter the measureds' dissembling and opacity with its own or (b) evoke neither from the measured. The social sciences cannot assuredly and ethically achieve (b) in overt measuring unless it is

measurers' and measureds' cooperative measuring. Because (a) is patently unethical for open democratic societies' sciences, its employment by them requires justification in the public interest in order to be ethically feasible.

Devices such as temporally separated and multiply differently put questions or test items dealing with the same matter can be used for trying to detect measureds' dissembling or for evading their opacity, but the measured must not recognize them as such devices (see, e.g., Irvine, 2010; Wellman, 1964). Response-set items have been used to try to detect measureds' dissembling (e.g., Berg, 1967; Block, 1965; Krause, 1965; Nicholson & Hogan, 1990; Walsh, 1990). Indirectness, as in projective tests (e.g., Sundberg, 1977, pp. 201-226), about which Meyer et al. (2001) reported: "...assessments with the Rorschach and TAT do not produce consistently lower validity coefficients than alternative personality tests"), has been used for trying to avoid measureds' intentional or unintentional ("unconscious") dissembling. Such devices involve forms of measurer dissembling that if suspected by the measured may invite their dissembling or opacity.

It is possible for some interviewees and test takers to become sophisticated enough to recognize and elude measurers' dissembling by, for example, keeping track of what answers one has given to prior questions or test items and upon noticing that a plausible next answer may be inconsistent with a prior one taking care to avoid this inconsistency, by learning to recognize response set detection items and how to respond to them to evade detection of one's response set, by learning how responses to projective tests are interpreted and so how to respond to evoke interpretations one prefers... Therefore, overt social science measurement must continually develop counters to the probably increasing sophistication of measureds who prefer to not be erroneously or disadvantageously measured (see Krause, 2018, on this). It is this preference itself that must ultimately be addressed, so that the measured come to prefer to be validly measured because they see openness to be in their own interest and so see themselves and the measurers as allies rather than adversaries (see, e.g., McCarthy, Hrabluik & Jelley, 2009) in the measuring. All of which clearly is beyond the scope of physical science "hard system" measuring.

Possible Consequences of Measuring and Measurements on the Measured

Simply to overtly psychologically measure persons can somewhat change them by, for example, their learning from the process of the *measuring* itself something they had not realized about themselves (perhaps their ignorance of chemical formulae or their having some Alzheimer's signs or symptoms) or their becoming something they had not

been (such as suspicious about what a test is actually measuring or interested in learning more about a topic they were tested on). From the *measurements* obtained (such as scoring unexpectedly high on a mental illness or intelligence assessment) they can become aware of important possibilities (Markus & Nurius, 1986) such as needing psychotherapy or getting into an elite college on a scholarship (e.g., Bobba & Frisancho, 2016; Fireman, Kose & Solomon, 2003; Godin, Sheeran, Conner & Germain, 2008; Kluger & DeNisi, 1996; Ohlsson, 1990). Social science still knows too little about the variety of consequences for the measured of (a) expecting to be, being, or realizing that one had been measured on some social science dimension, (b) being informed or left uninformed about what their measurements were, and (c) the consequences for those unmeasured who identify with (or abhor) those measured of being informed, for example, about the latter's intelligence, sexist, racist, or patriotism measurements.

Some persons are at least in part selected for employment, military service, education, promotion, parole, etc. on the basis of their T as inferred from their biographical data or psychometric test scores. Some persons may at least in part judge the nature and quality of their own lives on the basis of such measurements and rely on these for deciding how they will or ought to henceforth live their lives. Although social science has no monopoly on the measuring of T, it is responsible for obtaining valid and useful measurements of these because it has the obligation as a public enterprise to validly and usefully describe how persons live their lives, explain why so, predict how they will do so, and inform efforts to optimize the long run distribution of how well human lives are lived. In open democratic societies this means measuring overtly with the cooperation of the measured and informing of their measurements those who choose to be informed.

Insofar as everyone is not open to being measured on some dimension and insofar as valid measurements on this dimension are systematically different for persons who are opaque or dissembling than for persons who are open about this dimension, accurate generalization from measurements on samples of one such class of persons to the other class of persons cannot be achieved. Some undetected dissemblers are included in probably every sample of persons measured on any social science dimension and so somewhat bias the obtained measurement distribution on the dimension. Opacity, however, is obvious because it entails missing data.

Missing data from samples of persons measured suggest which of these persons may be wholly or partially opaque, although some missing data will have other explanations. The causes of missing data need careful study, because the persons whose missing data these are (and those like them among the meant to be generalized to unmeasured) may

systematically differ from the open measured (and those like them in this regard among the to be generalized to unmeasured) and bias the measured's measurement distributions on some dimensions and so the generalizing of these to the unmeasured. Accurate imputation of missing psychological data is impossible to prove and so not demonstrably achievable by statistical exercises in generalizing from obtained to missing data in order to facilitate full sample size data analysis (Allison, 2009; Di Franco, 2014; Little & Rubin, 2002). Instead, the social sciences must strive to avoid opacity as well as duplicity of the measured by making social science measuring a cooperative interpersonal craft, something physical science measurement has no need to do, is not concerned about, and so in this regard too cannot provide a proper model for psychological measurement.

Conclusions

Concern about the impact of their measuring and measurements on persons measured and on those unmeasured who know of the measuring or its findings is necessary for the social sciences. This measuring and publication of its findings are *interventions* on the measured and on some of the unmeasured who learn about the measuring or measurements. These interventions can influence what valid subsequent similar measurements would be on some of the measured and unmeasured. Thus, psychological measurements cannot be reliably valid *representations* of the measured unless they cooperate in obtaining valid measurements on themselves (see Hacking, 1983, on the notions of intervention and representation). The social psychology of social science measuring is in its infancy, but its crucialness for such measuring is what ultimately distinguishes social science from physical science measuring.

References

- Allison, P. D. (2009) Missing data. (pp. 72-89). In R. E. Millsap & A. Maydure-Olivares (Eds.). *The Sage handbook of quantitative methods in psychology*. Thousand Oaks, CA: Sage.
- Berg, I. A. (Ed.) (1967). *Response set in personality assessment*. Chicago, IL: Aldine.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*, 317-335.
- Block, J. (1965). *The challenge of response sets*. New York, NY: Appleton-Century-Crofts.
- Bobba, M., & Frisancho, V. (2016). *Learning about oneself*. Inter-American Development Bank Discussion Paper No. IDB-DP-450.

SOFT SYSTEM MEASUREMENT

- Bolino, M. C., Kacmar, K. M., Turnley, W. H., & Gilstrap, J. B. (2008). A multi-level review of impression management motives and behaviors. *Journal of Management, 34*, 1080-1109.
- Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment, 23*, 532-544.
- Bracha, H. S., & Burkle, F. M. (2006). Utility of fear severity and individual resilience scoring as a surge capacity, triage management tool during large-scale, bio-event disasters. *Prehospital and Disaster Medicine, 21*, 290-296.
- Broen, W. E., & Wirt, R. D. (1958). Varieties of response sets. *Journal of Consulting Psychology, 22*, 237-240.
- Burkle, F. M. (1996). Acute-phase mental health consequences of disasters: Implications for triage and emergency medical services. *Annals of Emergency Medicine, 28*, 119-128.
- Cano, S. J., Vosk, T., Pendrill, L. R. & Stenner, A. J. (2016). On trial: The compatibility of measurement in the physical and social sciences. *Journal of Physics: Conference Series 772*, 012025.
- Carter, N. T., Daniels, M. A. & Zickar, M. J. (2013). Projective testing: Historical foundations and uses for human resources management. *Human Resource Management Review 23*, 205–218.
- Chan, D. (2014). So why ask me? Are self report data really that bad? (pp. 309-336). In C. E. Lance & R. J. Vandenberg (Eds.). *More statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*. New York, NY: Routledge.
- Chang, H. & Cartwright, N. (2008). Measurement. (pp. 367-375). In S. Psillos & M. Curd (Eds.). *The Routledge companion to philosophy of science*. New York, NY: Routledge.
- Checkland, P. (2000). Soft systems methodology: A thirty year retrospective. *Systems Research, 17*, S11–S58.
- Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. New York, NY: Routledge.
- Cizek, G. J., Rosenberg, S. L. & Koons, H. H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement, 68*, 397–412.
- Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly, 3*, 307-331.
- Cornett, J. & Knight, J. (2009). Research on coaching. (pp. 192-216). In J. Knight (Ed.). *Coaching: Approaches and perspectives*. Thousand Oaks, CA: Corwin.
- Cozby, P. C. (1973). Self-disclosure: A literature review. *Psychological Bulletin, 79*, 73-91.
- Crocco, A. G., Villasis-Keever, M., & Jadad, A. R. (2002). Analysis of cases of harm associated with use of health information on the internet. *Journal of the American Medical Association, 287*, 2869-2871.
- Danziger, K. (1985). The methodological imperative in psychology. *Philosophy of the Social Sciences, 15*, 1-13.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford.

- DeMaio, T. J. (1984). Social desirability and survey measurement. (pp. 257-282). In C. F. Turner & E. Martin (Eds.). *Surveying subjective phenomena*, 2. New York, NY: Russell Sage.
- Di Franco, G. (2014). An alternative procedure for imputing missing data based on principal components analysis. *Quality & Quantity*, 48, 1149-1163.
- Dilchert, S., Ones, D. S., Viswesvaran, C. & Deller, J. (2006). Response distortion in personality measurement: Born to deceive, yet capable of providing valid self-assessments? *Psychology Science*, 48, 209-225.
- Edwards, A. L. (1990). Construct validity and social desirability. *American Psychologist*, 45, 287-289.
- Ellis, A. P., West, B. J., Ryan, A. M., & DeShon, R. P. (2002). The use of impression management tactics in structured interviews: A function of question type? *Journal of Applied Psychology*, 87, 1200-1208.
- Finkelstein, L. (2005). Problems of measurement in soft systems. *Measurement*, 38, 267-274.
- Fireman, G., Kose, G., & Solomon, M. J. (2003). Self-observation and learning: The effect of watching oneself on problem solving performance. *Cognitive Development*, 18, 339-354.
- Fiske, D. W. (1971). *Measuring the concepts of personality*. Chicago, IL: Aldine.
- Gibby, R. E. & Zickar, M. J. (2008). A history of the early days of personality testing in American industry: An obsession with adjustment. *History of Psychology*, 11, 164-184.
- Glaser, B. G., & Strauss, A. L. (1964). Awareness contexts and social interaction. *American Sociological Review*, 669-679.
- Godin, G., Sheeran, P., Conner, M., & Germain, M. (2008). Asking questions changes behavior: Mere measurement effects on frequency of blood donation. *Health Psychology*, 27, 179-184.
- Goffman, E. (1978). *The presentation of self in everyday life*. New York, NY: Penguin Random House.
- Griffith, R. L., & Peterson, M. H. (Eds.). (2006). *A closer examination of applicant faking behavior*. Greenwich, CN: Information Age Publishing.
- Habermas, T. (2007). How to tell a life: The development of the cultural concept of biography. *Journal of Cognition and Development*, 8, 1-31.
- Hacking, I. (1983). *Representing and intervening*, New York, NY: Cambridge University Press.
- Harano, R. M., Peck, R. C., & McBride, R. S. (1975). The prediction of accident liability through biographical data and psychometric tests. *Journal of Safety Research*, 7, 16-52.
- Hewitt, P. L., Flett, G. L., Sherry, S. B., Habke, M., Parkin, M., Lam, R. W., ... & Stein, M. B. (2003). The interpersonal expression of perfection: Perfectionistic self-presentation and psychological distress. *Journal of personality and social psychology*, 84, 1303-1325.
- Hoffmann, B. (1964). *The tyranny of testing*. New York, NY: Collier.
- Hui, C. H., & Triandis, H. C. (1985). The instability of response sets. *Public Opinion Quarterly*, 49, 253-260.
- Human Performance. (2011). Uncovering the Nature of Applicant Faking Behavior: A Presentation of Theoretical Perspectives. *Human Performance*, 24.

- Hyman, H. H., & Singer, E. (Eds.). (1968). *Readings in reference group theory and research*. New York, NY: Free Press.
- Irvine, S. H. (2010). The foundations of item generation for mass testing. (pp. xxxi-xxxiii & 3-14). In S. H. Irvine & P. C. Kyllonen. (Eds.). *Item generation for test development*. New York, NY: Routledge.
- Kinsman, J. (2012). "A time of fear": Local, national, and international responses to a large Ebola outbreak in Uganda. *Globalization and Health*, 8, 15.
- Kluger, A. N. & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.
- Krause, M. S. (1965). Role-deviant respondent sets and resulting bias: Their detection and control in the survey interview. *Journal of Social Psychology*, 67, 163-82.
- Krause, M. S. (2012). Measurement validity is fundamentally a matter of definition, not correlation. *Review of General Psychology*, 16, 391-400.
- Krause, M. S. (2013). The data analytic implications of human psychology's dimensions being ordinally scaled. *Review of General Psychology*, 17, 318-325.
- Krause, M. S. (2016). Case sampling for psychotherapy practice, theory, and policy guidance: Qualities and quantities. *Psychotherapy Research*, 26, 530-544.
- Krause, M. S. (2017). Item response theory requires logically unjustifiable assumptions. *Quality & Quantity*, 51, 1549-61.
- Krause, M. S. (2018). The scientific study of the qualities of individual human lives, rather than of their average quantities in aggregations of lives. *Quality & Quantity*, 28, 58-75.
- Krause, M. S. (submitted). The nature and functions of psychological theory.
- Krause, M. S., Lutz, W., & Böhnke, J. R. (2011). The role of sampling in clinical trial design. *Psychotherapy Research*, 21, 243-251.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, 47, 2025-2047.
- Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107, 34-47.
- Lee, R. M. (1993). *Doing research on sensitive topics*. Thousand Oaks, CA: Sage.
- Lewin, K. (1951). *Field theory in social science*. New York, NY: Harper.
- Little, R. J. A. & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: Wiley.
- Luce, R. D. & Narens, L. (2008) Theory of measurement. (pp. 523-33). In I. Blume & S. N. Durlauf (Eds.) *Palgrave Dictionary of Economics* (2nd ed). New York, NY: Springer.
- Luft, J. (1961). The Johari window: A graphic model of awareness in interpersonal relations. *NTL Human Relations Training News*, 5, 6-7.
- Mari, L. (2005). The problem of foundations of measurement. *Measurement*, 38, 259-266.
- Mari, L. (2013). A quest for the definition of measurement. *Measurement*, 46, 2889-2895.

- Mari, L., Carbone, P., Giordani, A., & Petri, D. (2017). A structural interpretation of measurement and some related epistemological issues. *Studies in History and Philosophy of Science Part A*, *65*, 46-56.
- Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, *41*, 954-969.
- Marsh, (1984). Do polls affect what people think? (pp. 565-592). In C. F. Turner & E. Martin (Eds.). *Surveying subjective phenomena*, *2*. New York, NY: Russell Sage.
- Maul, A., Irribarra, D. T., & Wilson, M. (2016). On the philosophical foundations of psychological measurement. *Measurement*, *79*, 311-320.
- Maul, A., Mari, L., Irribarra, D. T., & Wilson, M. (2018). The quality of measurement results in terms of the structural features of the measurement process. *Measurement*, *116*, 611-620.
- McCarthy, J., Hrabluk, C. & Jelley, B. (2009). Progression through the ranks: Assessing employee reactions to high-stakes employment testing. *Personnel Psychology*, *62*, 793-832.
- Meier, D. & Wood, G. (Eds.) (2004). *Many children left behind*. Boston, MA: Beacon Press.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., ... & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, *56*, 128-165.
- Michell, J. (2008). Is Psychometrics pathological science? *Measurement: Interdisciplinary Research and Perspectives*, *6*, 7-24.
- Michell, J. (2010). The quantity/quality interchange: A blind spot on the highway of science. (pp. 45-68). In J. Valsiner & A. Toomela (Eds.). *Methodological thinking in psychology: 60 years gone astray?* Charlotte, NC: Information Age Publishing.
- Miller, C. E., & Barrett, G. V. (2008). The coach-ability and fake-ability of personality-based selection tests used for police selection. *Public Personnel Management*, *37*, 339-351.
- Monfils, L. F., Firestone, W. A., Hicks, J. E., Martinez, A. C., Schorr, R. Y., & Camilli, G. (2004). Teaching to the test. (pp. 37-62). In W. A. Firestone, R. Y. Schorr & L. F. Monfils (Eds.). *The ambiguity of teaching to the test: Standards, assessment, and educational reform*. Mahwah, NJ: Erlbaum.
- Morwitz, V. G., & Pluzinski, C. (1996). Do polls reflect opinions or do opinions reflect polls? The impact of political polling on voters' expectations, preferences, and behavior. *Journal of Consumer Research*, *23*, 53-67.
- Moy, P., & Rinke, E. M. (2012). Attitudinal and behavioral consequences of published opinion polls. (pp. 225-245). In C. Holtz-Bacha & J. Strömbäck (Eds.). *Opinion Polls and the Media*. London, UK: Palgrave Macmillan.
- Nichols, A. L. & Edlund, J. E. (2015). Practicing what we preach (and sometimes study): Methodological issues in experimental laboratory research. *Review of General Psychology*, *19*, 191-202.
- Nicholson, R. A. & Hogan, R. (1990). The construct validity of social desirability. *American Psychologist*, *45*, 290-292.
- Ohlsson, S. (1996). Learning from performance errors. *Psychological Review*, *103*, 241-262.

- Paulhus, D. L. (2001). Socially desirable responding: The evolution of a construct. (pp. 51-73). In H. I. Braun, D. N. Jackson, & Wiley, D. E. (Eds.). *The role of constructs in psychological and educational measurement*. New York, NY: Routledge.
- Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PLoS ONE*, *11*, e0146487. doi: 10.1371/ journal.pone.0146487
- Perse, E. M. (2001). *Media effects and society*. Mahwah, NJ: Erlbaum.
- Popham, W. J. (2001). Teaching to the test: High crime, misdemeanor, or just good instruction. *Educational Leadership*, *58*, 16-20.
- Posner, R.A. (1981). The economics of privacy. *The American Economic Review*, *71*, 405-409.
- Rossi, G. B & Crenna, F. (2016). Toward a formal theory of the measuring system. *Journal of Physics: Conference Series* 772.
- Rothschild, D., & Malhotra, N. (2014). Are public opinion polls self-fulfilling prophecies? *Research & Politics*, *1*, 1-10.
- Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High-stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, *63*, 215-227.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*, 93-105.
- Schwarz, N., Knäuper, B., Oyserman, D., Stich, C., Sirken, M., Hermann, D., Schechter, S. et al. (2008). The psychology of asking questions. *Evaluation*, *22*, 127-160.
- Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, *39*, 68-74.
- Shishkin, I. F. (2016). Measurements of non-physical quantities. *Journal of Physics: Conference Series* 772, 012029.
- Smith, H.J., Dinev, T. & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly*, *35*, 989-1016.
- Spencer-Oatey, H. (2007). Theories of identity and the analysis of face. *Journal of Pragmatics*, *39*, 639-656.
- Stark, S., Chernyshenko, O. S., Chan, K.-Y., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology*, *86*, 943-953.
- Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, *5*, 1-25.
- Sudman, S. & Bradburn, N. M. (1982). *Asking questions*. San Francisco, CA: Jossey-Bass.
- Sundberg, N. D. (1977). *Assessment of persons*. Englewood Cliffs, NJ: Prentice-Hall.
- Teghtsoonian, R. (1989). The study of individuals in psychophysical measurement. (pp. 95-102). In G. Ljunggren & S. Dornic (Eds.). *Psychophysics in action*. Berlin, DE: Springer.
- Teller, P. (2008). Representation in science. (pp. 434-441). In S. Psillos & M. Curd (Eds.). *The Routledge companion to philosophy of science*. New York, NY: Routledge.

- Tourangeau, R. & Bradburn, N. M. (2010). The psychology of survey response. (pp. 315-346). In P. V. Marsden & J. D. Wright (Eds.). *Handbook of survey research* (2nd ed.). Bingley, UK: Emerald Group Publishing Limited.
- Van Praag, B. (2011). Well-being inequality and reference groups: An agenda for new research. *The Journal of Economic Inequality*, 9, 111-127.
- Walsh, J. A. (1990). Comments on social desirability. *American Psychologist*, 45, 289-290.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L. B., & Grove, J. B. (1981). *Nonreactive measures in the social sciences*. (2nd ed.) Boston, MA: Houghton Mifflin.
- Wellman, F. L. (1964). *The art of cross examination*. (4th ed.). New York, NY: Simon and Schuster.
- Wetzel, E., Böhnke, J. R., & Brown, A. (2016). Response biases. (pp. 349-363). F. T. L. Leong, D. Bartram, F. Cheung, K. F. Geisinger & D. Iliescu (Ed.). *The ITC international handbook of testing and assessment*. New York, NY: Oxford University Press.
- Wilson, M. (2013). Using the concept of a measurement system to characterize measurement models used in psychometrics. *Measurement*, 46, 3766-3774.
- Zickar, M. J. & Broadfoot, A. A. (2014). The partial revival of a dead horse? Comparing classical test theory and item response theory. (pp. 37-60). In C. E. Lance & R. J. Vandenberg (Eds.). *More statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*. New York, NY: Routledge.
- Zickar, M. J. & Gibby, R. E. (2006). A history of faking and social desirability responding on personality tests. In R. L. Griffith & M. H. Peterson (Eds.). *A closer examination of applicant faking behavior*. Charlotte, NC: Information Age Publishing.
- Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement*, 20, 71-87.
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology*, 84, 551-563.
- Ziegler, M., MacCann, C., & Roberts, R. (2011). Faking: Knowns, unknowns, and points of contention. (pp. 3-18). In M. Ziegler, C. MacCann & R. Roberts (Eds.). *New perspectives on faking in personality assessment*. New York, NY: Oxford University Press.