



## THE CONCEPTUAL VALIDITY OF EMPIRICAL SCALE CONSTRUCTION: THE CASE OF THE SENSATION SEEKING SCALE

Jeremy S. H. Jackson\* and Michael Maraun

Department of Psychology, Simon Fraser University, Burnaby, B.C., Canada V5A 1S6

(Received 25 May 1995; received for publication 6 December 1995)

**Summary**—An assessment of the present state of sensation seeking (SS) research suggests that the concepts of physical risk taking and sensation seeking have become synonymous with version five of Zuckerman's Sensation Seeking Scale (SSS). Given this, it would appear that the validity of SS research rests on the extent to which the SSS actually measures risk taking and sensation seeking tendencies. The present paper suggests that a series of logical and conceptual errors in the development of the SSS have resulted in a scale that does not measure sensation seeking. This raises concerns about both the value and meaning of investigations in which it is employed. It is concluded that risk taking and SS research has learned more about the SSS than sensation seeking itself. Copyright © 1996 Elsevier Science Ltd.

### INTRODUCTION

Since the introduction of Zuckerman's Sensation Seeking Scale (SSS), considerable research interest has been directed toward risk taking and sensation seeking behavior. Studies have assessed numerous personality, lifestyle, biological, cognitive, attitudinal, psychopathological and demographic correlates of sensation seeking (Zuckerman, 1979). In the vast majority of these studies, the SSS is the standard measure of the hypothetical sensation seeking trait. Bromiley and Curley (1992) state that the term sensation seeking has "narrowed to be almost synonymous with measures of the Sensation Seeking Scale" (p. 94). From this it follows that sensation seeking research has produced empirical associations, conclusions and theories that are about individuals' SSS scores. Therefore, the contribution of this research to our knowledge of sensation seeking proper depends on whether the SSS actually measures sensation seeking.

In this paper we argue that the use of purely statistical methods in the development of the SSS has resulted in a set of SSS items that do not measure the construct of sensation seeking. This claim is based on the argument that empirical correlations between items do not provide any justification for claims about their conceptual relationship (i.e. that the items have similar meaning). Since the choice of SSS items was based solely on empirical item correlations there is no justification for the claim that SSS items have similar conceptual meaning. An example involving traffic exposure and economic data is used to show how this misapplication of empirical methodology can produce scale items that have strikingly little correspondence to the constructs they are designed to measure. The example shows that unless conceptual clarification of the meanings of scale items is undertaken *prior* to the study of empirical relationships, there is little justification for claims to the effect that research centering on these items is relevant to the construct of interest. A content analysis of SSS items demonstrates that, in fact, this failure to precede an empirical analysis with a conceptual one has resulted in the inclusion of items that are not conceptually linked to the SS construct. Finally, we respond to the claim that conceptual clarity is not necessary because research with the existing SSS items is likely to have produced empirical relationships that are close to what would have been found with a conceptually justified measure of sensation seeking.

The arguments contained in this paper represent a general critique of the currently popular brand of measurement theory outlined by Cronbach and Meehl (1955). In our experience, psychologists are not generally aware that Cronbach and Meehl (1955) give only one of many treatments of the measurement problem—the justification for the claim: ' $X$  is a measure of  $Y$ '. Consequently, many

\* To whom all correspondence should be addressed.

psychologists do not realize that there might be other treatments of the measurement problem that undermine the logical coherence of the techniques invented by Cronbach and Meehl. Some other brands of measurement theory include axiomatic measurement theory, the latent variable tradition and Guttman's facet theory (Krantz, 1991). In part, Guttman's facet theory can be seen as an attempt to deal with a problem that is not satisfactorily addressed by Cronbach and Meehl—the fundamental distinction between conceptual and empirical issues. In this paper, we explore this distinction and the ramifications it has for psychological test development in general and the SSS in particular. We begin with a short description of the development and characteristics of the SSS.

### THE SENSATION SEEKING SCALE

Zuckerman (1979) originally developed the Sensation Seeking Scale (SSS) in order to assess the tendency of individuals to differ with respect to their optimal levels of stimulation and arousal. The original 54 items chosen for Form I of the scale consisted of 14 items pertaining to preference for extremes of sensation, 12 items relating to dangerous activities, 2 need for general excitement items, 4 need for adventure items, 8 items expressing preferences for the new and unfamiliar, 8 preference for irregularity items and 6 items expressing a desire for exciting (as opposed to reliable and predictable) friends. In later work with Form III and Form IV of the SSS, 63 new items were included. The new items reflected the dimensions assessed by the previous set but also included items relating to attitudes toward sex (Zuckerman, 1971). These 113 items (4 were excluded due to extreme endorsement–non-endorsement proportions) represented the item pool used in the construction of the most recent version of the scale, Form V.

Form V of the SSS consists of 40 forced choice items, each of which belongs to one of four 10-item subscales. Items were chosen for each subscale by factor analyzing the items used in the previous form of the scale (Zuckerman, Eysenck & Eysenck, 1978). The 10 items with the highest factor loadings for each of the first four factors extracted constituted the four subscales. Although some items did not meet the predetermined minimum factor loading criterion of 0.3, all of the items correlated most highly with the factor they were chosen to represent. On the basis of item content, the four factors were named Thrill and Adventure Seeking (TAS), Experience Seeking (ES), Disinhibition (DIS), and Boredom Susceptibility (BS).

Although the initial SSS items were chosen for their agreement with Zuckerman's concept of sensation seeking, the selection of subscale items was made solely on the basis of statistical criteria. That is, subscales were defined as the first four factors extracted in a factor analysis of the original 113 items, while individual subscale items were chosen on the basis of their factor correlations. The following argument suggests that this reliance on statistical/empirical rather than conceptual item selection criteria is inappropriate. In a continuation of this point, it is demonstrated in a conceptual analysis that the SSS items do not assess the concepts for which they are named. We begin by showing that empirical relationships between items do not provide grounds for claiming that the items involved have similar conceptual meaning.

### THE EMPIRICAL CONCEPTUAL DISTINCTION

It is generally accepted by practitioners of social science that a high correlation between variables indicates that they measure the same thing. This belief is reflected in the reasoning of Zuckerman (1979) when he questions the validity of three of the Form IV subscales: "The correlations between some of the factor scales (ES, DIS, and BS) were too high; although we expected some correlation between factors...the magnitude of the correlation indicated too much overlap and redundancy" (p. 105). In an attempt to reduce the Form IV subscale redundancy, a series of correlational analyses were conducted in order to select a new set of items for the Form V subscales. In a comparison of the Form IV and Form V items, Zuckerman comments on the adjustments required by the statistical procedures: "The most drastic change was on the ES scale, where three new items (IV: 11, 15, and 22) had to be included. One (IV 25) came from the Form IV BS scale, whereas the other two had not been in any of the previous factor scales" (p. 106). The choice of a procedure in which statistical criteria determined the selection of both full scale and subscale items (and even the switching of

items from one subscale to another) means that Zuckerman's claim that a given item *is*, for example, a BS item is based solely on its web of empirical relationships with other items. However, if it were not true that large empirical item correlations indicate that items have similar conceptual meaning (i.e. are about the same thing), then this scale construction strategy would not provide support for the claim that the SSS items are about sensation seeking.

In a succinct statement of the relationship between correlation and conceptual meaning, Guttman (1977) claims that "correlation does not imply content" (p. 99). He reasons that if a correlation did imply the content of the two items involved, then the answer to the question, 'what is variable  $x$  about?', could always be determined by examining the content of a variable that has a medium or high correlation with  $x$ . In a closely related point, Baker and Hacker (1982) argue that since the empirical relationship between variables is no indication of their conceptual relationship, statistical evidence can never serve as the criterion for establishing the meaning of a psychological construct. In an illustration of this point, Guttman (1977) suggests that the support for the statement '2 + 2 = ? is a mathematical ability item' is not in its intercorrelations with other items, but in the conceptual *fact* that it *is* a mathematical question. What makes the item '2 + 2 = ?' appropriate for a test of mathematics is the very fact that this item has *meaning* as a mathematics item. This state of affairs rests on a conceptual standard of correctness: one is correct if one calls this a mathematics item and incorrect if one calls this, e.g. a geography item. The item has a normative role in the practice of mathematics, and as such, no set of correlations with other non-mathematical items could have bearing on its meaning. In fact, since this item may correlate more highly with items that assess nonmathematical constructs such as verbal ability or trivia knowledge than with mathematical ability itself, reliance on *empirical* relationships could actually lead to incorrect conclusions about what the item measures.

To illustrate the bankruptcy of empirically based attempts to establish meaning we present a fictitious scenario centering on a set of recently observed correlations between quarterly economic factors, traffic exposure rates and a mystery variable named ' $X$ ', for the province of British Columbia between 1981 and 1991. For the purpose of this illustration, the identity of variable ' $X$ ' will be temporarily withheld. In Table 1 it can be seen that  $X$  has very large, positive correlations with 'number of registered vehicles', 'number of kilometers driven', and 'average auto insurance premium cost'.

Following Zuckerman's strategy, we should be able to pin down the meaning of  $X$  in terms of the items that are highly correlated with it. According to this strategy of definition by correlation,  $X$  must correspond to the intersection (i.e. the nexus) of these three variables, and so we will take the *meaning* of  $X$  to be defined nomologically as 'general driving activity' (GDA). Further nomological investigation reveals, however, that GDA has small correlations with several economic variables (see Table 1), and so we refine the meaning of  $X$  to be 'non-economically based general driving activity' (NGDA).

With the meaning of  $X$  soundly established, research is carried out on the empirical aspects of  $X$ , known as NGDA. Correlations are computed between  $X$  and other variables, means are calculated, and little by little an empirical picture of  $X$  is built. The rapidly accumulating results are then used in the construction of elaborate theoretical models designed to account for  $X$ . This scientific program is then declared to have brought us considerably further in our understanding of non-economically based general driving activity. There is, unfortunately, one major problem;  $X$  is actually the provincial gross domestic product! Thus, it follows that since the research conclusions and theories are

Table 1. Simple Pearson correlations between the economic and driving exposure variables, and the unknown factor  $X$

| Driving exposure and economic variables | Unknown factor $X$ |
|---|--------------------|
| Registered vehicles                     | 0.81*              |
| Kilometers driven                       | 0.83*              |
| Average auto insurance premium cost     | 0.92*              |
| Unemployment                            | -0.29              |
| Inflation rate                          | 0.14               |
| Disposable income                       | 0.23               |

\* $P < 0.05$ .

about driving exposure and the data they are based upon are about the gross domestic product, the research conclusions are both groundless and meaningless.

The forms of this example are analogous to the construction of, and research program based on the SSS. In both cases, empirical item correlations were taken as saying something fundamental about the meaning of the items involved. In the example, this practice resulted in a false characterization of the meaning of the mystery variable *X*. In the SSS, this practice may also have resulted in a similarly false characterization of the meaning of the items chosen to measure SS. The example also shows that research about a concept that is based on a set of items that do not have meaning for the concept produces groundless and meaningless results. But since the same practice of definition by correlation was employed in both the example and the SSS, there is the distinct possibility that research results using the SSS may also be both groundless and meaningless.

This illustration demonstrates that since the meanings or contents of items cannot be established on the basis of empirical results, empirical results can not provide justification for claims about what the items measure. Baker and Hacker (1982) comment on the consequences of carrying out empirical work in the absence of prior conceptual clarification.

“Unless the experimentalist has first clarified the psychological concepts that he employs in conceptualizing the data he wishes to explain or explore, unless he separates out the conceptual puzzles (which are legion) from the empirical questions, his experiments can be guaranteed to multiply confusion” (p. 229)

In the example, confusion resulted because the meaning of *X* was not determined prior to the analyses of empirical results. In this paper we have argued that the empirical methodology adopted by Zuckerman has projected the same sort of confusion into the SSS.

In the selection of items to measure a concept it is the meaning of the items that is at issue. Since the empirical relationship between items is irrelevant to issues of meaning, the only justification for claims about the meaning of items is conceptual (Maraun, 1989). Baker and Hacker (1982) suggest that conceptual clarification can only be obtained from:

“...the description of the use of mental expressions, of the circumstances in which they are employed, the complex grammatical structures in which they occur (and those in which they cannot significantly occur), of the behavior in different circumstances which provides grounds for their use, and of the purposes and roles of the utterances in which they occur” (p. 230).

Thus, the validity of SSS items as measures of SS rests not on their intercorrelations with other factors, but on the extent to which they instantiate (are denoted by) the concept of SS. It is important to note that since empirical scale construction techniques do not provide grounds for conceptual claims, the conceptual ‘validity’ of scale items assumes considerably more importance. This is because, given the irrelevance of empirical evidence, the *only* justification for the inclusion of an item on a given scale is conceptual.

We have shown that what Zuckerman owed to the quest for an SS measurement instrument was a coherent explication of the meaning of *sensation seeker*, and, on the basis of this explication, a set of items that are denoted by the concept. In the following section, a conceptual analysis of the concept of sensation seeking shows that in fact the SSS does not measure sensation seeking. The support for this claim will come from our addressing the fundamental definitional issues neglected by Zuckerman. This will be attempted in a description of a number of the logical contours of the correct (and incorrect) uses of the concept of sensation seeker.

#### CONCEPTUAL CLARIFICATION AND SENSATION SEEKING

Despite the fact that the term *sensation seeker* rings somewhat like a cliché, it has a number of obvious features. Firstly, *sensation seeker* is a dispositional term because its instantiation rests on behavioral criteria. That is, for the term to be applied to an individual he/she must have actually pursued, and engaged in, sensational activities. It makes no sense to say, for example, that ‘John is a sensation seeker, but has never sought sensation’. Secondly, the concept has an open-circumstance relativity (Baker & Hacker, 1982). What this means is that, given the correct background circum-

stances, practically any 'raw' behavior *could* instantiate the concept. Thus, there does not exist a finite set of behaviors called sensation seeking behaviors, rather, widely ramifying conditions of employment.

The grounds of application of a concept are often usefully analyzed by explicating its relations to other concepts. The term sensation seeker has strong ties to the concepts of intentionality and purpose. Bill is not a sensation seeker if he, e.g. (a) accidentally parachutes; or (b) parachutes against his will. Bill must have intended to engage in such an activity, since the notion of intent is a precondition of instantiation that is clear from the meaning of the term 'seeker' (i.e. someone who actively looks for something). Furthermore, behavior that instantiates sensation seeker is conceptualized as being for the purpose, or sake, of sensation itself. Hence, a bridge painter who walks out onto the girders of a bridge 300 feet above a gorge is not a sensation seeker, while Sue, who does the same thing for 'the thrill of it', is a sensation seeker.

We will now bring this brief analysis to bear on the SSS. Before turning to an analysis of individual items, we will document several serious problems at the structural level. Zuckerman states that he chose the initial 113 items to represent the eight sensation seeking categories outlined previously. Although this initial step might seem to some to be akin to conceptual work, his particular choice of categories manifests a decided lack of conceptual clarity. This is because an initial conceptual clarification of the sort given here would have disqualified items that do not instantiate a disposition, such as 'needs', 'preferences' and 'attitudes'. An individual can have the strongest need to be a sensation seeker imaginable without actually being one (Alston, 1975). Likewise, a preference is not relevant since an individual can always have a preference for *X* without choosing, seeking or pursuing *X*. A sensation seeker does not *prefer* sensational experiences, but instead actively pursues them, and in fact, does have them.

The second structural point is that there is no justification for the four subscale motif of the SSS. This is because the decision that SS is conceptually four dimensional was based purely on the analysis of inter-item correlations. But we have shown that empirical item correlations are irrelevant to issues concerning the meaning of sensation seeking. In fact, as we have shown, only the meaning of the concept can determine whether it incorporates a conceptual dimensional distinction. But since only empirical results were relevant in the investigation of the dimensionality of the concept, there is no justification for claiming that the dimensions uncovered by the empirical results are meaningful.

At the item level there are two fundamental problems. The first is the large number of non-criterial items. That is, items that do not provide logical justification for the application of the predicate sensation seeker to an individual who endorses them. Many of these items can be thought of as conceptual question marks since a more detailed specification of their background conditions would be required to link them to any concept. Others refer to 'wishes', 'preferences' and 'desires', and so cannot instantiate the dispositional term sensation seeker. Consider the item 'I often wish I could be a mountain climber'. Setting aside the issue of whether mountain climbing is necessarily a sensational experience, tautologically to *wish* to engage in this activity is not the same as to engage in it. But the dispositional term sensation seeker is instantiated by behavior, not wishes or desires.

The second problem is a direct consequence of the employment of correlational techniques in the context of scale construction. Specifically, correlational techniques invite the scale developer to make the logical mistake of confusing a criterion with a correlate. The height of mountain is an empirical correlate of the depth of snow on the mountain. It would be incorrect, however, to take *depth of snow* as a criterion for the *height of mountain*, because the concepts have different meanings. It may be the case, for example, that sensation seekers typically prefer 'clashing colors' and thus would endorse the item 'I often find beauty in the 'clashing' colors and irregular forms of modern painting'. However, this preference does not provide grounds for the application of the concept, being at best an empirical correlate of being a sensation seeker. The mistake perhaps rests on a common scale construction exercise: The investigator envisions a sensation seeker and lists behaviors in which he/she is likely to engage. This is a logically dubious practice, according to which we might well conclude that what it is to be a president is to travel in bullet proof vehicles.

The content analysis undertaken here reveals that even though there are *empirical* relationships among the items of the SSS, the majority of items do not pertain to SS. However, it is often held that even if a measure is not conceptually what it should be, it can still be of use if it produces

*empirical* results that are similar to those that would have been found with a conceptually justifiable measure. We present two responses to this idea. First, since it is clear that the empirical similarity of variables is no guarantee of their conceptual similarity, the only way to establish the empirical justifiability of a given measure is to actually compare the results it obtains to those found with a conceptually justifiable measure. However, if one had access to a conceptually justifiable measure, there would be little reason to retain the questionable measure.

The second response is illustrated in the following example. Consider a conceptually justifiable measure of sensation seeking,  $y$ , and a questionable measure,  $x$ . Suppose that although  $x$  and  $y$  are not conceptually equivalent, they have a strong positive correlation,  $r_{xy}=0.65$ , and that  $x$  has a correlation of 0.6 with a criterion variable  $z$ . The following expression provides bounds for the correlation between  $y$  and  $z$ :

$$r_{xy}r_{yz} - (1 - r_{xy}^2)(1 - r_{yz}^2) < r_{yz} < r_{xy}r_{yz} + (1 - r_{xy}^2)(1 - r_{yz}^2)$$

Given that  $r_{xy}=0.65$  and  $r_{xz}=0.6$  (which incidentally are as large as the highest inter-item/item-factor correlations found by Zuckerman),  $r_{yz}$  is bounded by  $-0.22$  and  $1$ . Thus, even if the conceptually questionable measure,  $x$ , is strongly empirically related to the conceptually justifiable measure,  $y$ , it is by no means guaranteed that  $x$  will produce similar empirical results.

#### RESPONSES TO SOME COMMON OBJECTIONS

##### *Objection:*

“How do we know that conceptual analyses will lead to correct conclusions about what an item measures?”

*Response:* This position does not attempt to provide a framework for determining beyond doubt when *particular* measurement claims are correct or incorrect. The attempt is to support the contention that the proper justification for the claim that ‘ $X$  is a measure of  $Y$ ’ can only be based on conceptual/logical argument. This is because empirical correlations can not guarantee anything about the meaning of the items involved. Logical/conceptual argument, however, speaks directly to the meaning of the items in question. For example, what justifies that a person has taken a measurement of height is that they have followed the rules for taking a measurement of height. The rules are logically internal to the concept of height. This can be seen in how an individual goes about justifying that a measurement of height has been taken. They do not show that the value they have obtained is associated with other aspects of the measured object. They simply point to the rules that are constitutive for the measurement of height and make the case that their actions are in conformity with the rules.

##### *Objection:*

“The SS scales have generated testable deductions that have been experimentally verified.”

*Response:* It is never correct to form ‘testable deductions’ about the meaning of items, or whether a particular item measures a particular construct. This is because ‘testable deductions’ refer to propositions that stand in need of empirical support. As we have shown, empirical results can never provide justification for measurement claims.

##### *Objection:*

“Through empirical research with the SSS we will gradually discover the true nature of sensation seeking.”

*Response:* Coherent scientific investigation must distinguish between discoveries about a particular phenomenon and what constitutes a measurement of that phenomenon. This is because an empirical discovery about  $X$  is predicated on a correct measure of  $X$ . If one can not justify that the initial measurement was taken of  $X$ , then it follows that one can not justify that the empirical discovery is about  $X$ . Therefore, if it can not be established prior to empirical work that a particular measurement,

$X$ , is of sensation seeking, then it can not be justified that the empirical results involving  $X$  are relevant to sensation seeking. This means that the claim that research involving  $X$  will enable us to gradually discover the true nature of sensation seeking can not be justified unless it can be determined prior to empirical investigation that  $X$  is a measure of sensation seeking. As we have shown, the justification for a measurement claim rests on conceptual/logical argument, not empirical evidence.

## CONCLUSION

The argument made in this paper can be summarized as follows:

1. One cannot support a measurement claim with items that do not pertain to the concept of interest. For example, there is no justification for the use of the item  $2+2=?$  as a measure of physical fitness.
2. Empirical item correlations can not establish whether items pertain to a given concept. For example, the fact that a verbal ability item has a higher correlation with the item  $2+2=?$  than a mathematical ability item does not establish that  $2+2=?$  is a verbal ability item.
3. Zuckerman employed empirical item correlations in an attempt to determine which 40 items pertained to the concept of SS.
4. Therefore, there is no logical justification for the claim that the empirically chosen SSS items are measures of SS.
5. In a conceptual analysis, it was shown that Zuckerman's treatment does, in fact, do damage to the meaning of SS, and that many of the items chosen to measure SS do not pertain to the concept of SS.
6. Therefore, the SSS is not a measure of SS.
7. Empirical results are relevant to a particular concept only if they are based on a measure that pertains to the concept.
8. Therefore, empirical results obtained with the SSS are not relevant to the concept of SS.

This argument shows that the requirement in bringing a set of items together to measure a concept is that they do, in fact, pertain to the concept (i.e. their meaning is given by the concept). As shown in the traffic exposure example, empirical correlations cannot establish whether items do, in fact, pertain to the same concept. The sole use of empirical results in the construction of a measurement scale can therefore lead to the inclusion of items that have no meaning for the concept of interest. The analytic form that does have the power to justify the claim that a given item is a sensation seeking item is conceptual analysis. Zuckerman, however, neglected conceptual clarification in the development of the SSS. As we have shown, this neglect has resulted in the selection of SSS items that do not measure the concept of sensation seeking. Furthermore, our analysis shows that even if there were a way of knowing that there existed a strong empirical relationship between the SSS and a conceptually correct measure, the claim that little has been lost in using the SSS as a 'stand in' measure of the actual concept lacks justification. It seems then, that until a conceptually valid measure of sensation seeking is constructed and employed, SS research will remain shrouded in ambiguity.

## REFERENCES

- Alston, W. (1975). Traits consistency and conceptual alternatives for personality theory. *Theory of Social Behavior*, 5, 17-48.
- Baker, G. P. & Hacker, P. M. S. (1982). The grammar of psychology: Wittgenstein's Bemerkungen Uber Die Philosophie Der Psychologie. *Language and Communication*, 2, 227-244.
- Bromiley, P. & Curley, S. P. (1992). Individual differences in risk taking. In Yates, J. F. (Ed.) *Risk taking behavior*. Chichester: Wiley.
- Cronbach, L. J. & Meehl, P. E. (1995). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Guttman, G. G. (1977). What is not what in statistics. *Statistician*, 26, 81-107.
- Krantz, D. (1991). From indices to mappings: The representational approach to measurement. In Brown, D. & Smith, J. (Eds) *Frontiers of mathematical psychology: Essays in honour of Klyde Coombs*. New York: Springer-Verlag.
- Maraun, M. (1989). Conceptual and methodological problems in the act frequency approach. Paper presented at the annual meeting of the Canadian Psychological Association, Halifax.

- Zuckerman, M., Eysenck, S. & Eysenck, H. J. (1978). Sensation seeking in England and America: Cross-cultural, age and sex comparisons. *Journal of Consulting and Clinical Psychology, 46*, 139–149.
- Zuckerman, M. (1971). Dimensions of sensation seeking. *Journal of Consulting and Clinical Psychology, 36*, 45–52.
- Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level of arousal*. Hillsdale, NJ: Erlbaum.