An Empirical Attempt at Evaluating Stress: A Failure Discovered Through Cross-Validation

J. P. Dobbin
Department of Psychology
Western University

Edward Helmes
Department of Psychology
James Cook University

Rod A. Martin
Department of Psychology
Western University

Abstract
Stress remains popular as a psychological construct. Different aspects of stress are emphasized depending upon the environmental issue, target population, and measure used. Existing measures are often confounded between causes of stress and effects of stress and also may emphasize a particular perspective on stress. Here we evaluate the empirical method of item selection as an alternative for developing a stress scale, using salivary cortisol levels as the empirical criterion. Items were adapted from measures of perceived stress, daily hassles, and life events as used in two studies of stress that measured salivary cortisol. Correlations with cortisol levels led to the retention of 75 items of the pool of 535, which were administered to a third sample of 28 medical students. The 75-item scale did not correlate with cortisol levels. Of 15 individual items that did, six correlated in the opposite direction to that predicted. Results illustrate the dangers of empirical item selection methods.

Key words: Stress; Cortisol; Test Construction; Perceived stress; Daily hassles; Life events

INTRODUCTION
With the continued wide use and popularity of the term, there is increasing acceptance of the term "stress". Selye [1] provides a history of the term, noting that it involved a short- or long-term taxing environmental event or any perceived threatening event [2, 3]. Either of these may affect any number of the emotional, physiological, or biochemical system [4, 5]. A general modern definition is provided by Morris [6] for usage in psychology: “1. Specifically, a physical or emotional reaction to a situation perceived as unfamiliar, threatening, harmful, and so on. 2. The negative situation itself.” (p. 2110).

Given the variability inherent in the definition, it is not surprising that there are a variety of ways of measuring stress by self-report [7]. Many of these are questionnaires, and some simply quantify the number of major life events that a person has experienced, while others focus on minor life experiences and "daily hassles" (e.g., [8, 9]. Still others ask about emotional or other subjective reactions to work [[10]. The items on the majority of these scales seem high in face validity, in that all seem to be related to the construct of stress. The scales often have high internal consistency, and show significant correlations with criterion measures, such as physical illness and psychiatric symptomatology. However, often there is not a clear
demonstration that the criterion measures of illness or symptomatology are in fact related to stress, or whether the degree to which an event is uncontrollable, unpredictable, threatening, or taxing is, in fact, stressful. This makes the construction of measures using the atheoretical empirical methods of scale construction as typified by the widely used Minnesota Multiphasic Personality Inventory [11] more attractive.

The early measures of stress were based upon major life changes that occurred over a specified period of time [12]. Later studies refined elements of the stress reaction, such as time urgency [13] and well-being [14]. A simple listing of stressful life experiences does not take into account the actual impact of the events upon the person’s life. Holmes and Rahe [15] developed a questionnaire which measured both the number and the severity of life events occurring. The events were chosen because they required a degree of adaptation (e.g., a change of marital status, employment, or death of a close friend) [16].

Despite its face validity, life event research has been criticized [4]. One of these criticisms was that some events on the inventories appear related to the criteria they were used to predict. For example, someone with a slowly developing illness may become less productive at work, receive pressure to improve, lose contact with friends, and become a burden upon a spouse as disease processes develop. When the illness is diagnosed, adjustment problems in the person’s life may be construed as having caused the illness.

In 1966, Lazarus [2] proposed a model of psychological stress which accounted for individuals not only in the appraisal of potentially threatening conditions, but also in the appraisal of the person’s abilities to cope. This model suggested that there were a number of ways in which individual differences could modulate responses to a threatening environment. Lazarus thus proposed that stress be defined as an interaction between individuals and their environment. A central tenet is that people differ in their beliefs, motives, education, abilities, and social resources, and that such differences play an important role in determining whether or not an experience will be construed as threatening. The Lazarus [2, 3] model led to the development of the Perceived Stress Scale (PSS) as a measure of "appraised stress" [17, 18]. The fourteen PSS items ask the respondent to estimate, on a scale ranging from "never" to "very often", how often events have been perceived as stressful, anxiety provoking, overwhelming, and uncontrollable. Again, the PSS is reported to be a significant predictor of a number of indices of health, including depression scores, frequency of visits to health clinics, physical symptoms, and adherence to a smoking reduction program. However, Lazarus, Delongis, Folkman, and Gruen [19] have argued that the PSS overlaps events antecedent and consequent to a stressful event. Individuals who obtain elevated scores on criterion measures which assess emotional or physiological symptoms may very well report feeling distressed, nervous, and upset that things were not going well. Thus, a significant association between scores on the PSS and scores on criterion measures may be inflated simply due to the amount of content variance shared. Such an association is less likely to occur if items are selected on empirical grounds.

In contrast to stressful events, Kanner, et al. [8] proposed that the relatively minor daily irritants which most people experience represent an important source of stress that is not evaluated in other measures. Such "daily hassles" refer to experiences such as being stuck in traffic or annoyed by smokers. Under the argument that stress results as a result of the accumulation of these minor events, Kanner et al demonstrated that minor daily irritants account for a greater proportion of variance in psychological symptomatology than do major life events. Martin [20] noted that daily hassles account for more variance in mood disturbance than do major life events, even after the variance due to major events has been
parceled out. In contrast, if hassles are first parceled out, major life events do not add significantly to the variance accounted for in mood disturbance. Interestingly, the daily hassles scale has been criticized for the same reasons that Lazarus and others have criticized other stress scales. There were significant confounds between the daily hassles scale and other measures [21].

Since the problem of confounded measures seems to be an inherent property of scales intended to measure stress, is it time to head back to the theoretical blackboard? Perhaps there should be two separate research endeavors: first determine whether the construct of stress exists independently of illness, and then second, test to see if there is any association between the construct stress and the construct illness.

**PHYSIOLOGY OF STRESS**

The early models of stress [22] emphasize the physiological reactions in response to threatening events [23]. The activation of the sympathetic nervous system and hypothalamic-pituitary-adrenal (HPA) axis was central to this physiological reaction [24]. The production and release of adrenal hormones, including cortisol, produce a variety of physiological changes, including increased heart rate and respiration, greater blood flow to the muscles and brain, and changes in the metabolism of serum proteins and sugars. These physiological changes have well-recognized survival advantages for animals that are threatened in that they prepare the animal for greater amounts of activities so that it may either combat the threat or run away.

Although much early research examined these processes in animals, many results have been found to be applicable to humans, particularly associated with the onset of depression [25]. As they are highly reliable, they may form an ideal criterion for determining whether an individual has recently experienced threat as reflected in excitation of the sympathetic division of the autonomic nervous system, known to be a common response to challenging or threatening experiences in humans [26, 27]. Such overt physiological symptoms may form the core of items that reflect empirical indicators of stress.

The secretion rate of stress hormones increases significantly during autonomic arousal and while cortisol may be measured in a serum, only 5% of serum cortisol level is biologically active. The free fraction of cortisol is associated with levels in saliva and urine [28]. This method has been shown to be diagnostic in both normal and clinical populations [29, 30, 31].

The use of cortisol measured from saliva appears to be a reliable and valid biochemical measure of responsivity to a variety of events that could reasonable seen as taxing, challenging and threatening. Given the complexity and cost of obtaining this measure, an alternate inexpensive method, such as a questionnaire, would be highly desirable. The present paper thus describes an attempt to develop a paper and pencil measure of stress, as defined by changes in cortisol values produced in response to an event. The scale was developed empirically, by selecting items on the basis of their association with cortisol levels. This method has a long history within psychological assessment, although it has been subject to criticism [32]. We attempted to compensate for the most important of these, the capitalization upon chance in performing repeated statistical tests, by cross-validating the selection items in a different sample.
METHOD – STAGE 1

Selecting Items
Items for the scale were derived from a series of studies in which subjects completed questionnaires and provided saliva samples for the quantification of salivary cortisol. The questionnaires included the Perceived Stress Scale [17], the Daily Hassles Scale [8], the Telic Dominance Scale [33], Profile of Mood States [34], and the Life Events of College Students [35]. In both studies, subjects also completed other scales not reported here.

Cortisol was quantified from the saliva samples by RIA [31]. Materials were obtained from Nuclear Diagnostics, and each assay was run in duplicate with known control values to test the reliability of the assay. The intra-assay correlation was .91, and control samples fell within the expected ranges.

Participants
Participants in both studies were introductory psychology students, participating in return for course credit. In the first study, 42 subjects (19 males and 32 females) produced saliva samples and completed questionnaires on two occasions: once immediately after spring break during a period with relatively few academic demands, and once during final exam week. Each individual subject participated at the same time of day on both occasions. In the second study, 30 subjects (14 males and 16 females) provided saliva samples and completed questionnaires for five consecutive days.

Procedure
Correlations were calculated between cortisol levels and individual items from each of the questionnaires. Because items from different questionnaires had different numbers of response options, items from different scales would produce unequal variances across items. This problem was resolved by transforming all items to standard scores. In order to control for the collection of data at different points of time during the day, and thus at different points during the daily cycle of cortisol levels, time of day was partialled out of the correlations of items with cortisol levels.

A total of 535 items was evaluated. At a Type I error level of .05, it was expected that 27 items would be selected on the basis of chance alone. The analysis was repeated in both samples.

RESULTS – STAGE 1

Forty items were selected from the first study, with an internal consistency (co-efficient alpha) of .85. Forty-two items selected in the same manner from the second study led to a co-efficient alpha of .70.

After reviewing the content of the items, seven were excluded on the basis that they were nearly identical to another item. This led to a total of 75 items chosen for the first draft of the questionnaire. The 75 items used are reproduced in the Appendix. Because the majority of response scales was on a five point Likert scale, the remaining items were rescaled to this format as well. The majority of the items showed a plausible association between “stress” and item content.

Cross Validation
The 75 item questionnaire was appended to a battery of questionnaires in a separate research project in which cortisol levels were also evaluated.
METHOD – STAGE 2

Participants
Twenty-eight first-year medical students, with a mean age of 23.4 years, were tested during their first week back to school after spring break.

Materials and Procedure
A 30 ml sample of heparinized blood was obtained from the antecubital vein of each subject when they arrived for the study. They were then given a 10 x 17 mm test tube, and asked to provide 5 ml of saliva. They then completed the battery of questionnaires. Total cortisol levels of plasma were quantified in addition to cortisol levels from saliva. Plasma was extracted from whole blood by centrifuging samples at 300 g for 20 minutes. Aliquots of plasma were placed in fresh tubes and frozen at -70 degree C until assayed. Materials for assay of cortisol were obtained from Calastead Laboratories. Cortisol levels were assayed from both plasma and saliva using materials with the same lot numbers.

RESULTS – STAGE 2
The 75 item scale gave a co-efficient alpha of .88, and with the deletion of five items with low-item total correlations, alpha rose to .90.

The intra-assay correlation was .98 for plasma cortisol and .88 for salivary cortisol. The control samples fell within the expected ranges for both plasma and salivary cortisol. The mean plasma cortisol level was 145.26 N MOL-L, while the mean salivary cortisol level was 8.64 N MOL-L. Both values fall within the normal expected range for samples. As would be expected, total cortisol levels of plasma were unrelated to unbound cortisol assayed from saliva (r = .09, ns).

There was no association between scores on the stress scale and cortisol levels assayed from either plasma or saliva. When correlations between individual items and cortisol levels were evaluated, only 15 of the 75 items have a significant correlation. Furthermore, 6 of these 15 items were now correlated in the opposite direction to that initially obtained. For example, subjects who agreed with the item "I have been concerned about the health of a family member" apparently have lower levels of cortisol in the second study than people who were not worried (r = -.39, <.05).

GENERAL DISCUSSION
The purpose of this investigation was to select questionnaire items on the basis of their empirical association with cortisol levels, and then to cross validate those items in order to eliminate the less stable ones. The scale showed a high co-efficient alpha in both studies, suggesting a consistency in measuring the relevant construct. However, the total scores proved to be unrelated to cortisol levels (cf. [27]. Furthermore, of the original 75 items, only 15 retained a significant association with cortisol upon cross validation. In addition, a significant proportion of those 15 items showed a reversal in the direction of association.

With this degree of shrinkage upon cross-validation, it would appear that any further attempt at refinement of an empirically derived questionnaire would be futile.

Nevertheless, the results of this study raise several interesting issues regarding the nature of cortisol secretion during threat, and about self-report measures of responses to threat. The first issue is whether or not people are consciously aware of any physiological phenomena that may be associated with responses to threat. For example, sympathetic activation typically

URL: http://dx.doi.org/10.14738/assrj.28.1401.
includes increased heart rate, increased respiration, muscle tension, shakiness and warm hands. Such items appear to be theoretically sound, having been demonstrated to be associated with responses to threat, and may be salient enough for subjects to recall accurately whether they have recently experienced such symptoms of arousal.

One of the problems with asking subjects about these physiological indices of arousal is that many events may produce them. Such items may be endorsed by a subject who had just attended a fitness class, as well as by one who had just broken up with a lover. Thus, it would seem that in order to obtain a measure of response to threat, it may be necessary to develop questions regarding the subject's appraisal of recent events. Thus, a fitness class would be rated as less distressing than ending a relationship.

These results also confirm that the concept of “stress” is indeed a slippery one that varies over time within individuals dynamically [36, 37]. It would appear that perhaps the best method would be to return to a theoretical reconceptualization, and to derive measures based from that theory. Given the intuitive acceptance of the Lazarus [3] model, it would appear that this would be a fruitful starting point. Given the results found here, which confirm previous criticisms of empirical test development [32], it would appear that such a theoretical approach would be far more likely to be productive than an empirically based one.

**APPENDIX**

**JD Scale**
On the following pages is a list of 75 statements. Indicate the degree to which you agree or disagree with each statement, using the following scale:

A = strongly disagree  
B = disagree  
C = neutral  
D = agree  
E = strongly agree

1. I would prefer taking an evening course for fun over taking an evening course to improve my qualifications.

2. I would prefer improving an athletic skill by playing the game over improving it through systematic practice.

3. I often like to do things "for kicks"

4. I would prefer to always take holidays in the same place over taking holidays in many different places.

5. I prefer a continual unexpectedness or surprise in my life over a steady routine.

6. After having a disagreement with a close friend in which both of us are shouting very loudly, I would feel anxious.
7. After having a disagreement with a close friend in which both of us are shouting very loudly, I would feel excited.  

8. After having a disagreement with a close friend in which both of us are shouting very loudly, I would feel stressed.  

9. If the teacher announced that she would hand back exams in order of grade, beginning with the highest mark in the class, and my name was one of the first to be called, I would find this quite amusing.  

A = strongly disagree  B = disagree  C = neutral  D = agree  E = strongly agree

10. I vary an awful lot from one situation to another in the extent to which I laugh or otherwise respond with humor (i.e., it depends a lot on who you are with, where you are, how you feel, etc.)

11. One of my most outstanding characteristics is that I can be amused and laugh in a wide variety of situations.

12. I have often found that my problems have been greatly reduced when I tried to find something funny in them.

13. I often miss the comical point in a situation where others catch on.

14. I have often felt that difficulties were piling up so high that I could not overcome them.

15. I have generally been feeling playful.

16. I have generally been feeling worried.

17. I have generally been feeling emotional.

18. I have generally been feeling worked up.

19. I have generally been feeling hyper.

20. I have generally been feeling distressed.

21. I have generally been feeling fidgety.

22. I have generally been feeling tense.
23. I have generally been feeling worn out.  
24. I have generally been feeling shaky.  
25. I have generally been feeling sad.  
26. I have generally been feeling discouraged.  
27. I have generally been feeling miserable.  
A = strongly disagree  B = disagree  C = neutral  D = agree  E = strongly agree  
28. I have generally been feeling exhausted.  
29. I have had troubling thoughts about my future.  
30. I have been concerned about the health of a family member.  
31. I have been concerned about owing money.  
32. I have been bothered recently by my use of alcohol.  
33. I have had problems with employers.  
34. I am having problems with my boyfriend/girlfriend (wife/husband). Circle "C" if not applicable.  
35. I have the feeling that my work is unchallenging.  
36. I feel that I do not have enough time to do the things that I need to do.  
37. I have recently felt pressured by a supervisor.  
38. I have recently been concerned about news events.  
39. I have recently written an exam which was stressful for me.  
40. I have little control over the way I can spend my time.  
41. In the last day, I have not dealt successfully with irritating life hassles.  
42. In the last day, I have not been coping effectively with changes that are occurring in my life.
| 43. | In the last day, I have felt tense and nervous. | A | B | C | D | E |
| 44. | I am never late for meetings. | A | B | C | D | E |
| A = strongly disagree | B = disagree | C = neutral | D = agree | E = strongly agree |
| 45. | I sometimes put words in peoples' mouths to speed up conversations. | A | B | C | D | E |
| 46. | I consider myself to be more relaxed and easy-going. | A | B | C | D | E |
| 47. | People look to me for leadership more often than they look to others. | A | B | C | D | E |
| 48. | Coworkers and friends would agree that I "live, eat, and breathe" my work. | A | B | C | D | E |
| 49. | It would bother me if other people experienced more success than I. | A | B | C | D | E |
| 50. | I like most of my fellow workers, and enjoy my work. | A | B | C | D | E |
| 51. | I rarely engage in two or more activities at the same time. | A | B | C | D | E |
| 52. | I believe that organizations work best when employees do not compete with each other. | A | B | C | D | E |
| 53. | I seldom take work home with me. | A | B | C | D | E |
| 54. | I usually show up early for work to prepare things. | A | B | C | D | E |
| 55. | I often compare my performance to that of my coworkers. | A | B | C | D | E |
| 56. | I have been bothered by misplacing things. | A | B | C | D | E |
| 57. | Smokers are annoying. | A | B | C | D | E |
| 58. | I am bothered by my physical appearance. | A | B | C | D | E |
| 59. | I have had troubling nightmares lately. | A | B | C | D | E |
| 60. | I would rather compile a short-dictionary for financial reward then write a short story for fun. | A | B | C | D | E |
| 61. | I would rather recount an incident accurately | A | B | C | D | E |
than exaggerate for effort.

A = strongly disagree    B = disagree    C = neutral    D = agree    E = strongly agree

62. I would rather have continuity in the place I live than having to move frequently.

63. I typically finish my work before going out to enjoy myself.

64. I would enjoy climbing a mountain.

65. I like to plan ahead rather than taking each day as it comes.

66. Before attempting a new task, I feel excited and challenged.

67. I have often found it difficult to live up to my parents’ expectations.

68. My parents usually gave me a lot of praise and encouragement.

69. I have had a lot of difficulty making close friends.

70. I have not been as successful in my life as I should have been.

71. I am my own worst critic.

72. I have generally been feeling bushed.

73. I have recently begun counseling.

74. I have been feeling generally annoyed lately.

75. I would rather win a game easily than play with scores very close.

References


**FOOTNOTES**

The data reported here were collected by JPD who also wrote the initial report as part of his studies. This paper is submitted in his memory, deceased on 1 November 2007 at age 47.