Rapport A 2012:1

A note on the assessment of performance-based self-esteem

Lennart Hallsten

Karolinska Institutet

Contents

1 Preface (in Swedish)

LUST-projektets huvudsakliga syfte är att studera stress och hälsa i samband med övergången mellan högskoleutbildning och arbetsliv. I publikationer från projektet har fokus riktats mot både utbildningsfrågor samt arbetslivsfrågor (e.g., Djordjevic, Rudman & Gustavsson, 2011; Hasson, Omne-Pontén & Gustavsson, 2007; Rudman, Djordjevic, Frögéli & Gustavsson, 2009). Denna rapport skiljer sig lite åt från tidigare publikationer och är en utvärdering av ett mätinstrument som ingått i studien.

Rapporten vänder sig till forskare, studenter samt övriga som söker instrument för mäta teoretiskt viktiga variabler förknippade med stress och ohälsa. Ett stort tack till alla studiedeltagare vars svar och engagemang har varit en förutsättning för genomförandet av studien och dess kvalitet. Projektet möjliggörs genom forskningsanslag från AFA försäkring.

Stockholm den 26 mars 2012

Petter Gustavsson vetenskaplig ledare för LUST-projektet

2 Introduction

A short scale to assess performance-based self-esteem (PBSE) has been developed (Hallsten, Josephson, & Torgén, 2005) and used in several studies. Although the scale has shown many desirable psychometric properties, results of confirmatory factor analysis (CFA) have indicated that the scale is not unidimensional as assumed. In the present study, an additional item was added to the original PBSE scale in an attempt to make it unidimensional.

PBSE refers to a vulnerable qualitative aspect of self-esteem, its contingency. Self-esteem is said to be contingent if self-esteem is dependent on meeting particular conditions and standards within or across various domains. For instance, an individual's self-esteem can be contingent on his or her occupational achievements or physical appearance. Contingent self-esteem has been assumed to increase during socialization processes and during difficult and stressful conditions (Deci & Ryan, 2002; Dweck, 1999). Most likely, PBSE as an instance of general contingent self-esteem has its origin in introjected external performance standards prevailing in socialization settings.

Contingent self-esteem is a relatively stable orientation that often results in negative consequences in terms of behaviors, cognitions or emotions. Crocker and Park (2004) maintained that contingent self-esteem fosters self-regulation strategies based on inauthentic self-validation goals that undermine learning, autonomy and health, and that striving for self-esteem entails a number of long-term costs. Similarly, Dykman (1998) has found that striving for self-esteem and self-validation, rather than for growth or learning, is predictive of later depression. Furthermore, in her social-cognitive model of motivation (Dweck, 1999), Dweck has described similar self-validation strategies among so called entity-oriented students. Instead of espousing learning or task-oriented goals, these students tend to adopt performance goals such as looking smart and winning positive judgments, which often produce anxiety and helplessness.

An interindividual notion of contingent self-esteem can profitably be captured by an overall score or by a single scale for general contingent self-esteem (Burwell &

Shirk, 2006; Neighbors, Larimer, Geisner, & Knee, 2004; Patrick, Neighbors, & Knee, 2004). A rather extensively used instrument in Sweden and Scandinavia for assessing general self-esteem contingency is a questionnaire for PBSE (Hallsten, et al., 2005). PBSE refers to a compelling motive or orientation to gain or maintain self-esteem through good performance in roles or arenas of importance for self-esteem. The PBSE questionnaire was initially created to capture burnout from an etiological process view (Hallsten, 1993) and to distinguish burnout and its consequences from a similar, yet different stress-related syndrome called wornout (Hallsten, 2005; Hallsten, Voss, Stark, & Josephson, 2011). The PBSE questionnaire has also been used outside the field of burnout. The reasons have been that strivings to appear clever, capable and competent have been assumed to increase in late modern societies and that such strivings more generally may contribute to psychological distress and poor health (Albertsen, Rugulies, Garde, & Burr, 2010; Dahlin, Joneborg, & Runeson, 2007; Innstrand, Langballe, Espnes, Aasland, & Falkum, 2010; Löve, Grimby-Ekman, Eklof, Hagberg, & Dellve, 2010). PBSE has also been found to increase in demanding socialization settings such as higher education (Hallsten, Rudman, & Gustavsson, 2012).

2.1 The performance-based self-esteem scale (the PBSE scale)

The original PBSE questionnaire includes items on cognitions related to general contingent self-esteem, such as contingency and imperative beliefs as well as ego-oriented motives without references to any specific domain (Hallsten, et al., 2005). The original PBSE questionnaire consists of the following four Likert-type items: with a response format from 1 (“Fully disagree”) to 5 (“Fully agree”). The scoring method to obtain a PBSE scale has been to compute the mean of the responses to these four items.

(a) “I think that I sometimes try to prove my worth by being competent”,
(b) “My self-esteem is far too dependent on my daily achievements”,
(c) “At times, I have to be better than others to be good enough myself”,
(d) “Occasionally I feel obsessed with accomplishing something of value”;

Hitherto, the PBSE scale has been used in studies with nationally representative samples and large occupational and student samples, jointly involving over 50,000

individuals. Longitudinal data over one to three years have been obtained from these samples. The internal consistency of PBSE scale has generally been over $\alpha = .80$, and its stability over one and three years has been found to be substantial, $r = .68$ and $r = .64$, respectively (Hallsten, 2005; Hallsten, et al., 2012). A national cohort of nurses showed a stability of $r = .55$ over seven years from their first year of education to their fifth year at work, while the stability from the first to the fifth year at work was $r = .69$. Furthermore, the scale has shown both convergent and discriminative validity. In relation to the dimensions of the Contingencies of Self-Worth scale (Crocker & Wolfe, 2001), the PBSE scale correlated most positively with the external dimensions of the Contingencies of Self-Worth scale, i.e. to competence, approval from others, and physical appearance (Lindblad, 2003).

However, when analyzed in isolation, the original PBSE scale does not seem to quite fulfill the criteria of unidimensionality as presumed. Such a deviation is not unique for the PBSE scale. Other well-known and assumingly unidimensional scales have also demonstrated deviations from unidimensionality, such as Rosenberg's Self-esteem scale (Aluja 2007, Greenberger 2003) and the General Health scale (GHQ-12) (Sconfienza 1998, Hankins 2008, Smith 2010). Nonetheless, when the factorial validity of the PBSE scale was tested using confirmatory factor analysis (CFA) of data from a Swedish population sample in the year 2000 (Hallsten, et al., 2005), the four PBSE items did not seem to be expressions of a single latent factor. Though the standardised factor loadings for the items were high, ranging from .72 to .84, some of the absolute and relative fit indices demonstrated bad fit to a unidimensional model (χ^2 (df = 2, N = 3981) = 177.00, $p < .001$; SRMR = .028; CFI = .98 ; NNFI = .94; RMSEA = .152). The errors of two items had to be allowed to correlate to get a good fit (χ^2 (df = 1, N = 3981) = 9.40; $p = .002$; SRMR = .006; CFI = 1.00; NNFI = .99; RMSEA = .046). This may indicate a need for improving the factorial validity of the PBSE scale. An option could be to reduce the PBSE scale to three items, as such a short scale always has a perfect fit to a unidimensional model. However, such a 3-item model is often of minor interest, since it is saturated with no variance and no degrees of freedom. At least four items are required to build a testable unidimensional scale. In the case of PBSE, this can be obtained by adding and deleting items to and from the original PBSE scale.

2.2 Aim

In the present study, the primary aim was to create a revised, short, reliable and unidimensional PBSE scale by using CFAs on data from two national cohorts of nurses. As a further aim was to keep the revised scale short, only one item was added to the original PBSE scale. Whether a valid unidimensional scale was obtained, it would be compared with the original PBSE scale in terms of means, standard deviations, and reliabilities and the correlation between the scales would be computed.

3 Method

3.1 Participants and items

The CFAs were based on data from two national cohorts of Swedish nurses, who recently had started their professional career and who participated in the LANE study (Rudman, Omne-Pontén, Wallin, & Gustavsson, 2010). The assessments of the original and modified PBSE scores were carried out twice for these cohorts in the years 2009 and 2010. In both years the nurses were presented with the original PBSE questionnaire plus one further item:

(e) "I nearly always try hard to appear proficient in the eyes of others"
(in Swedish: "Jag eftersträvar nästan jämt att framstå som skicklig i andras ögon").

In one of the nurse cohorts, called EX2004 whose members graduated in 2004, 1086 and 1023 nurses responded to all five items in the years 2009 and 2010, respectively. In the other cohort, named EX2006 consisting of nurses who graduated in 2006, 953 and 950 nurses fully responded. In both cohorts 90 percent were females among those who responded to all PBSE items. The mean age in 2009 was 37 years (SD = 7.4) in the EX2004 cohort and 36 years (SD =7.1) in the EX2006 cohort. For further information on the LANE study, see (Rudman, et al., 2010).

3.2 Analyses

CFA was used to verify the suitability of a unidimensional model for the PBSE items. As the analyses were conducted on data from the two cohorts at two occasions, the same PBSE models or item sets were examined among four different student samples.

A stepwise analytic strategy was followed. The first CFAs for each cohort and occasion were executed to substantiate the earlier finding that a unidimensional measurement model would not suit the original PBSE scale. This original PBSE scale was here called model "4org" and it comprised the four items "a", "b", "c", and "d". In the presumed second step, the model fit was to be examined for an extended PBSE scale with five items, called model "5", where the new item "e" had been

added to the four original items. If this extended scale would not prove to fulfill the criterion of unidimensionality for each cohort at each year, one item from the original scale would be deleted. The order of item deletion was based on the corrected item-total scale correlation, where the item with the lowest correlation would be removed.

The maximum likelihood method was used in the CFAs to examine the covariance matrices of the items. Four goodness-of-fit indices were applied in addition to the χ^2 test to evaluate the fit of data to the presumed unidimensional measurement model: the Standardized Root Mean Square Residual (SRMR), the Comparative Fit Index (CFI), the Non-Normed Fit Index (NNFI) and the Root Mean Square Error of Approximation (RMSEA). The CFI and NNFI should be >.95 and SRMR and RMSEA should be <.05 for good fit. All CFAs were carried out with LISREL 8.80.

4 Results

The results from the CFAs are presented in Table 1. The outcomes for the baseline item set with the four original PBSE items, named model 4org, showed in line with earlier studies a bad fit to a latent unidimensional model for each cohort and year, at least according to the RMSEA criterion (RMSEA > .15). When the new item "e" was added to the original scale, labeled model 5, the fit was somewhat better in all analyses but far from good (RMSEA > .10).

Table 1
Results from the CFAs of the original and modified PBSE scales based on the nurse cohort data

Cohort	Year	N	Model	df	χ^2	SRMR	CFI	NNFI	RMSEA
EX2004	2009	1086	4org	2	52.69	.025	.982	.946	.158
			5	5	79.89	.026	.984	.967	.118
			4-a	2	32.07	.017	.990	.970	.113
			4-d	2	3.21	.006	1.00	.999	.024
EX2004	2010	1023	4org	2	103.45	.038	.960	.880	.229
			5	5	123.77	.035	.971	.943	.154
			4-a	2	33.39	.019	.988	.964	.120
			4-d	2	5.20	.008	.999	.996	.040
EX2006	2009	953	4org	2	47.93	.029	.978	.935	.161
			5	5	55.60	.026	.985	.969	.106
			4-a	2	16.13	.016	.993	.980	.084
			4-d	2	4.41	.009	.999	.997	.036
EX2006	2010	950	4org	2	91.43	.040	.960	.879	.226
			5	5	131.15	.039	.963	.927	.164
			4-a	2	43.72	.024	.981	.943	.142
			4-d	2	13.59	.016	.995	.984	.078

As none of these two models demonstrated a satisfactory factorial validity, an item from the original scale was to be deleted. Reliability analysis comprising corrected item-total correlations indicated that the most endorsed item, item "a", and least endorsed item, item "d", had similar, comparatively low corrected item-total correlations (in the range r = .69 to .71). Consequently, two further models with four

items were analyzed; one where the most endorsed item “a” was deleted, called model “4-a”, and one where the least endorsed item “d” was deleted, labeled model “4-d”.

The results from the CFAs of these latter 4-item sets are also displayed in Table 1. While model 4-a exhibited a rather bad fit to a unidimensional model according to the RMSEA criterion, model 4-d showed a fairly good fit for both cohorts at both measurement occasions. The EX2006 cohort in 2010 was a possible exception, as the fit here was just barely acceptable according to RMSEA ($p < .08$). However, in relation to the other three fit criteria, the correspondence to a unidimensional model for this item set was quite satisfactory also for this cohort in 2010. It can be added that the standardised factor loadings for the four items in model 4-d were quite similar for the two cohorts in both years (for item “a”: around .76; item “b”: around .89; item “c”: around .82; item “e”: around .75). This item set might then be used as a revised unidimensional PBSE scale.

The relations between a revised PBSE scale based on model 4-d and the original PBSE scale, model 4 org, are described in Table 2 with regard to means, standard deviations, coefficient alpha and Pearson correlation. As a 3-item scale always will produce a perfect fit to a unidimensional, saturated, model, the original PBSE scale was also reduced to three items, where item “d” was deleted. This shortened scale, named model “3org”, and its relations with the original PBSE scale and with the revised scale are presented in Table 2.

As the least endorsed item had been deleted both in model 3org and 4-d, the means for the latter two models were higher than for the original PBSE scale (the standardised differences between the means in terms of Cohen’s d was in the range .05 to .14 for the revised PBSE scale in the four data sets, while Cohen’s d was around .15 for the shortened 3-item scale), whereas the standard deviations tended to be the same in the four samples. The reliabilities, expressed as Cronbach’s α□ □ were high for all scales and the correlations between the original PBSE scale and the shortened and revised scale were high in both cohorts and at both occasions, $r \geq .96$. These data can be compared with the ones from the Swedish population sample collected in 2000 (N = 4760), where the mean was 2.79 (SD = 1.11) for the original 4-item PBSE scale and 2.91 (SD = 1.15) for the shortened 3-item scale (Cohen’s d =

0.11) (Hallsten, et al., 2005). The correlation between the original and shortened PBSE scale was r = .97 in the population sample.

Table 2

Means, standard deviations, Cronbach's alpha and Pearson correlations for the original ("4org"), shortened ("3org") and revised unidimensional PBSE scale ("4-d").

Cohort	Year	n	Model	M	SD	α	r_{4org}
EX2004	2009	1094	4org	3.14	0.98	.88	-
			3org	3.31	1.00	.87	.98
			4-d	3.19	0.99	.89	.97
EX2004	2010	1037	4org	3.11	0.97	.88	-
			3org	3.28	1.00	.87	.97
			4-d	3.17	0.98	.90	.96
EX2006	2009	953	4org	3.11	0.97	.86	-
			3org	3.26	1.00	.85	.97
			4-d	3.24	0.96	.86	.96
EX2006	2010	950	4org	3.14	0.98	.86	-
			3org	3.30	1.00	.85	.97
			4-d	3.29	0.96	.87	.96

5 Discussion

In the present study, an attempt was made to create a short, reliable and unidimensional PBSE scale that would involve at least four items. This was done by performing CFAs on data from two national cohorts of Swedish nurses, who responded twice to an extended PBSE questionnaire one year apart. The results from the CFAs revealed that a scale consisting of three items from the original PBSE scale, items “a”, “b”, and “c”, and the added new item “e”, labelled model 4-d, together built a scale for PBSE, which fulfilled all criteria of unidimensionality. The fit deviances between the modified PBSE models (model 5, model 4-a, and model 4-d) in terms of SRMR, CFI and NNFI were not great, but the RMSEA criterion was sensitive to the disparities between the models.

When it is important to have a pure unidimensional PBSE scale, the new PBSE scale based on model 4-d might be a preferred alternative. If this option is utilized, it should be noticed that the mean level of PBSE tends to increase in comparison with the original PBSE scale by around 10 percent of a standard deviation. In regard to other measurement properties, the revised scale seems to be quite similar to the original one. The similarities of the standardised factor loadings over the cohorts and measurement occasions also suggest a metric invariance, implying that the different items were similarly interpreted by the nurses. If item “e” is not available, an alternative might be to utilize the shortened PBSE scale with three items. In many circumstances, however, it can be assumed that the original PBSE scale may be satisfactory, as it correlates very highly with the revised PBSE scale. An advantage for the original scale is that normative data are available from nationally representative samples.

Finally, some limitations of the present study should be recognized. First, the samples used for deriving and validating the revised PBSE scale were less than perfect. All participants belonged to the same occupation, a clear majority of them were females, and they were all recently graduated nurses (in 2004 and in 2006). In addition, the participants had earlier responded to the original PBSE scale as nursing students and as nurses before the assessments in 2009 and 2010. How these circumstances influenced the participants are difficult to know. Presumably, these

conditions had their greatest effects on the mean levels of PBSE. Since young and well-educated persons as well as females tend to have higher PBSE scores, the currently obtained mean levels were probably higher than if a nationally representative occupational sample had answered the extended PBSE questionnaire. More important, however, presumably is that the factorial validity of the revised PBSE scale was less affected than its absolute mean level. The base for this assumption is the outcome earlier seen for a nationally representative population sample that involved the original and shortened PBSE scales. This population sample and the present nurse cohorts displayed quite similar results for the original and shortened PBSE scales with regard to relative mean levels, reliabilities and correlation. The standardised factor loadings for the three first items in the original scale for the national population sample were also quite similar to the ones obtained for the nurse cohorts. This suggests that the factorial validity of the revised PBSE scale was not vastly biased by the present participants' occupation, gender and age.

Another weakness of the study was that a test of measurement invariance over the cohorts and measurement occasions were not carried out. Given the similarity of the cohorts, this was not deemed necessary, but this should preferably be accomplished in future validation studies. A possible additional drawback was that only one further item was added to the original scale in the analyses. Although the aim was to preserve the shortness of the scale to facilitate its inclusion in large surveys, a future PBSE scale might gain from being based on a larger item pool.

Some strengths of the study should also be acknowledged. The data sets were derived from national samples that were representative of an occupation in Sweden. Furthermore, the samples were rather large and the data were acquired at two occasions.

Finally, the premise of the study may raise an issue: Should the deviation from a unidimensional model be regarded as a shortcoming to be altered? The possibility remains that the original PBSE items in fact describe different severity or developmental levels of PBSE. For instance, the endorsement of item "d" can be a manifestation of a late outcome in the development the latent trait of PBSE, which few participants had reached or will reach. A variation in item and person severity may result in a multidimensional scale. Such cases may preferably be analyzed by

Guttman scaling or by item response theory (IRT), as illustrated in a study of a traditional burnout scale (Gustavsson, Hallsten, & Rudman, 2010). In the future, a corresponding analysis can be conducted with the original or extended PBSE scale.

6 References

Albertsen, K., Rugulies, R., Garde, A. H., & Burr, H. (2010). The effect of work environment and performance-based self-esteem on cognitive stress symptoms among Danish knowledge workers. *Scandinavian Journal of Public Health, 38 (suppl 3)*, 81-89.

Burwell, R., & Shirk, S. (2006). Self-processes in adolescent depression: the role of self-worth contingencies. *Journal of Research on Adolescence, 16*(3), 479-490.

Crocker, J., & Park, L. E. (2004). The costly pursuit of self-esteem. *Psychological Bulletin, 130*(3), 392-414.

Crocker, J., & Wolfe, C. T. (2001). Contingencies of self-worth. *Psychological Review, 108*(3), 593-623.

Dahlin, M. E., Joneborg, J., & Runeson, B. (2007). Performance-based self-esteem and burnout in a cross-sectional study of medical students. *Medical Teacher, 29*, 43-48.

Deci, E. L., & Ryan, R. M. (Eds.). (2002). *Handbook of self-determination research*. Rochester, NY, US: University of Rochester Press.

Djordjevic, A., Rudman, A., & Gustavsson, P. (2011). *235 röster om glappet: Sjuksköterskors reflektioner om övergången mellan utbildning och yrkesliv 2 år efter sjuksköterskeexamen* (No. B 2011:1). Stockholm: Karolinska Institutet.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. New York, NY: Psychology Press.

Dykman, B. M. (1998). Integrating cognitive and motivational factors in depression: Initial tests of a goal-orientation approach. *Journal of Personality and Social Psychology, 74*(1), 139-158.

Gustavsson, J. P., Hallsten, L., & Rudman, A. (2010). Early career burnout among nurses: Modelling a hypothesized process using an item response approach. *Int J Nurs Stud, 47*(7), 864-875.

Hallsten, L. (1993). Burning out: A framework. In W. Schaufeli, C. Maslach & T. Marek (Eds.), *Professional burnout: Recent developments in theory and research* (pp. 95-113). Washington: Taylor & Francis.

Hallsten, L. (2005). Burnout and wornout: concepts and data from a national survey. In A.-S. G. Antoniou & C. L. Cooper (Eds.), *Research companion to organizational health psychology* (pp. 516-536). Cheltenham: Elgar Publ.

Hallsten, L., Josephson, M., & Torgén, M. (2005). *Performance-based self-esteem: a driving force in burnout processes and its assessment*. No 2005:4, Arbete & Hälsa 2005). Stockholm: National Institute for Working Life.

Hallsten, L., Rudman, A., & Gustavsson, J. P. (in press). Does contingent self-esteem increase during higher education? *Self and Identity, DOI:10.1080/15298868.2010.544872*.

Hallsten, L., Voss, M., Stark, S., & Josephson, M. (2011). Job burnout and job wornout as risk factors for long-term sickness absence. *Work, 38*(2), 181-192.

Hasson, D., Omne-Pontén, M., & Gustavsson, P. (2007). *Exit 2006: En landsomfattande populationsbaserad studentspegling av sjuksköterskeutbildningen* (No. B 2007:3). Stockholm: Karolinska Institutet.

Innstrand, S. T., Langballe, E. M., Espnes, G. A., Aasland, O. G., & Falkum, E. (2010). Personal vulnerability and work-home interaction: The effect of job

performance-based self-esteem on work/home conflict and facilitation. *Scand J Psychol, 51*(6), 480-487.

Lindblad, L. (2003). *Självkänsla och utbrändhet (Self-esteem and burnout).* Unpublished Psychologist thesis, Stockholms university, Stockholm.

Löve, J., Grimby-Ekman, A., Eklof, M., Hagberg, M., & Dellve, L. (2010). "Pushing oneself too hard": performance-based self-esteem as a predictor of sickness presenteeism among young adult women and men--a cohort study. *J Occup Environ Med, 52*(6), 603-609.

Neighbors, C., Larimer, M. E., Geisner, I. M., & Knee, C. (2004). Feeling controlled and drinking motives among college students: Contingent self-esteem as a mediator. *Self and Identity, 3*(3), 207-224.

Patrick, H., Neighbors, C., & Knee, C. R. (2004). Appearance-related social comparisons: the role of contingent self-esteem and self-perceptions of attractiveness. *Personality and Social Psychology Bulletin 30*(4), 501-514.

Rudman, A., Djordjevic, A., Frögeli, E., & Gustavsson, P. (2009). *Den nyfärdiga sjuksköterskans arbetsvillkor: En beskrivning av anställning, verksamhet och arbetsförhållanden 1 år efter utbildningen* (No. B 2009:1). Stockholm: Karolinska Institutet.

Rudman, A., Omne-Pontén, M., Wallin, L., & Gustavsson, J. P. (2010). Monitoring the newly qualified nurses in Sweden: the longitudinal analysis of nursing education (LANE) study. *Human Resources for Health, 8*(10), 1-17.

Karolinska
Institutet