

# Selection of questions to short-form versions of original psychometric instruments in MoBa\*

Kristian Tambs<sup>1</sup> and Espen Røysamb<sup>1,2</sup>

1) Norwegian Institute of Public Health, Division of Mental Health

2) University of Oslo, Department of Psychology

E-mail: kristian.tambs@fhi.no

## ABSTRACT

Original psychometric instruments are usually too lengthy and space-consuming to be suitable for general population based health studies. Usually, however, they can be abbreviated without losing more measurement precision than what can be accepted in such studies. Here we demonstrate that short-form versions of three instruments which are part of the MoBa study, and which include from one third to half the items in the original versions, correlate from 0.90 to 0.96 with the original version. This means that the short-form versions measure approximately the same characteristics as do the original instruments, and that they can safely be used for research purposes in MoBa.

This is an open access article distributed under the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Psychometric instruments used to measure psychosocial characteristics like behavior, personality, and mental health, usually consist of long lists of questionnaire items. There are thousands of established psychometric instruments available for such purposes, typically with from 10 to several hundred items. The instruments are made so long because each item just covers a small fraction of the sphere of the behavior, trait, or syndrome to be measured. Also the response to each single item will to a large extent reflect random fluctuation, implying that the law of large numbers requires several items with more or less the same meaning to eliminate some of the random errors. Population screening studies with a broad scope of health issues, such as the Norwegian Mother and Child Cohort Study (MoBa), usually do not have questionnaire space for such original instruments, implying that abbreviated instruments must be used. Such shortenings certainly affect the measurement precision, but often the precision remains sufficient for epidemiological purposes.

Sometimes copyright interests prohibit reducing the number of items in an established instrument. Whenever possible, long psychometric instruments have been abbreviated in MoBa. Sometimes only judgment based on theoretical or common sense consensus – or on previously published results on the covariance structure of the instrument – has informed the selection of items for short-form instruments. Often, however, already existing data materials with scores from the original instruments have been available. Such data can be used to empirically select the items which in combination give the score that best resembles the score from the full instrument.

### *Some examples of instruments that were abbreviated and used as short-form instruments in MoBa*

The SCL-25, designed to measure symptoms of anxiety and depression, was first described in 1980 (1). It is similar to, but not identical with the anxiety- and depression part of the SCL-90 (2). Ten items are designed to tap anxiety, the other 15 tap depression. Because of the relatively broad set of symptoms included in the SCL-25, it is often used as a measure of mental distress or global mental health. Short-form versions of the SCL-25 are included in the MoBa questionnaires at week 15 of pregnancy, week 30 of pregnancy, 6 months, 18 months, 36 months, 8 years and 13 years.

The Rosenberg Self-Esteem Scale (RSES) (Rosenberg, 1989) is probably the most commonly used instrument for assessing self-esteem. It consists of 10 items, shown in Table 1. A short-form RSES is included in the MoBa questionnaires at week 15 of pregnancy, week 30 of pregnancy, 6 months, 18 months and 36 months.

The General Self-Efficacy Scale (GSE) also includes 10 items, shown in Table 2. It was developed to measure optimistic self-beliefs about coping with a variety of difficult demands in life. The scale was originally developed by Matthias Jerusalem and Ralf Schwarzer in 1979 and has later been revised and adapted to many other languages (4-6). A short-form version of the GSE is included in the MoBa questionnaires week 30 of pregnancy and at 18 months.

The aim of this study was to report on the construction of short-form versions of the SCL, RSES and GSE, and to provide evidence for their reliability and validity.

\* Some of the results presented here have previously been reported in Norwegian (Tambs K (2004): Valg av spørsmål til kortversjoner av etablerte psykometriske instrumenter. Forslag til framgangsmåte og noen eksempler. I: Inger Sandanger, Knut Sørgaard, Guri Ingebrigtsen, Jan F. Nygaard (red.): *Ubevisst sjeleliv og bevisst samfunnsliv. Psykisk helse i en sammenheng. Festskrift til Tom Sørensen 60 års jubileum*. University of Oslo, 2004, 29 - 48, ISBN 82-92192-23-9) and are presented again with the permission of the copyright owner of the original publication.

**Table 1.** Rosenberg Self-Esteem Scale.

---

I feel that I am a person of worth, at least on an equal plane with others.  
 I feel that I have a number of good qualities.  
 All in all, I am inclined to feel that I am a failure.  
 I am able to do things as well as most other people.  
 I feel I do not have much to be proud of.  
 I take a positive attitude toward myself.  
 I am able to do things as well as most other people.  
 I wish I could have more respect for myself.  
 I certainly feel useless at times.  
 At times I think I am no good at all.

---

**Table 2.** The General Self-Efficacy Scale.

---

I can always manage to solve difficult problems if I try hard enough.  
 If someone opposes me, I can find the means and ways to get what I want.  
 It is easy for me to stick to my aims and accomplish my goals.  
 I am confident that I could deal efficiently with unexpected events.  
 Thanks to my resourcefulness, I know how to handle unforeseen situations.  
 I can solve most problems if I invest the necessary effort.  
 I can remain calm when facing difficulties because I can rely on my coping abilities.  
 When I am confronted with a problem, I can usually find several solutions.  
 If I am in trouble, I can usually think of a solution.  
 I can usually handle whatever comes my way.

---

Response categories:

1 = Not at all true 2 = Hardly true 3 = Moderately true 4 = Exactly true

## METHODS

### Data material

For an empirically based selection of the best combination of items in short form scales to be included in the MoBa, we needed access to already existing data materials with scores on the original instruments.

The SCL-25 data were available from the so-called Fourty Year Study, conducted by the National Health Screening Service, later part of the Norwegian Institute of Public Health, in some Norwegian counties during the nineteen eighties and nineties. The Fourty Year Study in Nord-Trøndelag county took place in 1989 and included the population aged 40-42 years and 65-67 years (<http://www.fhi.no/dokumenter/2319904f86.pdf>). Also, beyond the standard questionnaire content of the 40 year studies the study in Nord-Trøndelag included the SCL-25. A total of 8,806 persons were invited, of whom 6,380 subjects (72.5%) participated. The data material is described in more detail in a previous publication. There we demonstrated that a combination of only 5 items gives a sum score which correlates 0.92 with the sum score from SCL-25 (7). Two items, "Loss of sexual interest or pleasure" and "Thoughts of ending your life" were excluded from the HUNT questionnaire because they were believed to be perceived as offensive by some of the participants, leaving us with an incomplete "SCL-23" material. Blanks were substituted with sample mean values where less than 4 of the SCL items were missing. There were valid SCL scores for 5,999 subjects (2,993 males and 3,006 females).

The data material used to select items to a short-form RSES was made available by Mette Ystgaard (8).

It consists of data on RSES from 250 male and female adolescents aged 17-18 years in a normal population sample. The material was almost free from missing values, but in a few cases missing was recoded to sample mean values. A list of the items in the original RSES is shown in Table 1.

The Generalized Self-Efficacy Scale (GSE) was included in a questionnaire in the Sogn & Fjordane project conducted by the Norwegian Institute of Public Health in 1994. This population based study included 1583 respondents (aged 18), yielding a response rate of 63% (9). The original GSE contains ten items, shown in Table 2.

### Statistical method

We used stepwise multiple linear regression analysis to select the short-version items. The total (sum) score from the original instrument was used as the dependent variable and each single item as independent variables, entered stepwise one at the time. In this method the predictor that explains most of the variance is automatically selected in the first step, then the new predictor which gives the largest gain in explained variance is entered at each step. We used the SPSS default value  $p=0.1$  as criterion to remove a variable as predictor, which did not cause any predictors to be dropped from the model in the analyses described here.

Whereas the analyses of the RSES and GSE data were conducted straightforwardly, some modification had to be made with the SCL data. Firstly, some items, like faintness, dizziness or weakness and sleeping problems, are not good indicators of mental distress during pregnancy or soon after having given birth, but may just as well reflect states associated with the struggle of pregnancy or postnatal life. Secondly, we wanted as good measures as possible of anxiety and depression separately without sacrificing the measurement precision of the global scores. Finally, we wanted a balanced number of items tapping anxiety and depression. That required some interfering with the automatic stepwise procedure. In the first MoBa questionnaire the "SCL-5", already shown to be the optimal combination of five items (7), was chosen. For the next waves (from questionnaire at week 30 of pregnancy) the MoBa administration decided to psychometrically strengthen the SCL short-form version by adding three items. In a new wave of questionnaires, completed when the child is 13 years old, planned in the near future, another four depression items are planned to be added to the SCL-short-form scale. This 12-item version is so far planned to be completed only once in the MoBa study, so we will not include it in our analyses here, but we will report its reliability and its correlation with the full-scale instrument.

As a start of the analysis of the SCL data we entered the five items already shown to give the best solution and already used in the first MoBa questionnaire. The next items were then entered according to the criterion for stepwise inclusion, except when somatic symptoms

**Table 3.** SCL-25, anxiety and depression. Explained variance and correlation between the short form scores for each step in the linear regression analysis. One new item is added for each step.

Step #	Item added	Anxiety (A) or depression (D)	% explained variance of total score	Correlation between short-form score and score from original instrument			Alpha reliability for the short form instrument		
				Global distress	Depression	Anxiety	Global distress	Depression	Anxiety
1	22. Worrying too much about things	D	58	.76	.76				
2	4. Nervousness or shakiness inside	A	71	.84		.77	.74		
3	18. Feeling blue	D	78	.88	.85		.80	.74	
4	2. Feeling fearful	A	81	.90		.84	.83		.69
5	17. Feeling helpless about the future	D	84	.92	.89		.85	.80	
6	1. Suddenly scared for no reason	A	85	.92		.86	.86		.76
7	24. Feeling everything is an effort	D	88	.94	.92		.87	.83	
8	7. Feeling tense or keyed up	A	89	.94		.90	.88		.78

Alpha-reliability for full SCL-25 (23 items, two excluded): 0.91

believed to be highly related to pregnancy or to the postpartum period were automatically entered. In such a case, the unsuitable item was removed and the analysis rerun. At the same time we checked that the fit of the model was only marginally worse for the separate anxiety and depression scores than when fitting models for anxiety or depression separately. Also we checked that the model fit was only marginally worse than when allowing any of the SCL items (including those judged as unsuitable for pregnant women) to enter the model.

The value of the adjusted multiple R (the square root of the explained variance) will usually be slightly higher than the correlation between the sum of short-version item scores and the sum score from the original instrument. That is because the short-version sum scores are not weighted before summed. Weighting them with the regression coefficients would give exactly the same correlation value as the multiple R. We also calculated the intercorrelation between the short-form and the original scores. Typically, the intercorrelation is usually very close to the multiple R. In cases where it is clearly lower, it might be worthwhile weighting each short-form item with the regression coefficient before summing them. Cronbach alpha was also calculated.

For a further check, it is possible to examine the covariance structure of a psychometric instrument is unifactorial, using confirmatory factor analysis (CFA). Sometimes an instrument designed to measure only one dimension in fact include two or more dimensions. Such a multifactorial covariance structure in a device designed to measure a single dimension is considered to be a violation of the psychometric model assumption and to demonstrate poor structural validity. As an example, we examined one of the instruments, the GSE, both in its original and short-form version with a CFA to compare the model fit of the full and the abbreviated instrument. For this purpose we used the M-plus computer program, which produces a number of model fit indices, of which the RMSEA and the CFI are the ones most widely used.

Using MoBa data, we also observed the distribution of the short-form scales and the intercorrelations between the same short-form scores from various points of times.

## RESULTS

The results for SCL-25 are shown in Table 3. One single item, "Worrying too much about things", explains 58% of the variance for the global SCL-25 score, corresponding to the correlation value 0.76. Adding one item, "Nervousness or shakiness inside", giving a two-item "short-form SCL-2", increases the correlation to 0.84. An eight-item version gives the correlation 0.94 with the global score from the original instrument. The sum of the four depression items in the SCL-8 correlates at 0.92 with the depression score from the full SCL-25, and the sum of the four anxiety items correlates at 0.90 with the anxiety score from the SCL-25. The unweighted sum of the eight items gives almost the same correlation as does a weighted sum, expressed as the multiple R or square root of the explained variance,  $\sqrt{0.89} = 0.94$ . Cronbach alpha for the eight-item version was estimated to 0.88 for the global score, 0.83 for depression and 0.78 for anxiety. Corresponding values calculated with the MoBa data (questionnaires at week 15 of pregnancy, week 30 of pregnancy, 6 months, 18 months, 36 months) varied from 0.83 to 0.88 (mean value 0.85) for global distress, from 0.74 to 0.83 (mean=0.78) for depression, and from 0.74 to 0.77 (mean=0.75) for anxiety.

"Test-retest" correlations for SCL-8 global scores and for the anxiety and depression part, shown in Table 4, were observed with data from the Moba questionnaires completed during the period from the pregnancy week 30 till the child was three years old. The shortest time-lag between the observation times was 7 months (between questionnaires at week 30 of pregnancy and 6 months), the longest was just over three years (questionnaires at week 30 of pregnancy and 36 months).

As mentioned a 12 item SCL is planned to be completed by the mothers when the children are 13

**Table 4.** Correlations between SCL10 scores at different points of time.

	Global distress			Anxiety			Depression		
	Q4	Q5	Q6	Q4	Q5	Q6	Q4	Q5	Q6
Q3, pregnancy week 30	.53	.49	.46	.47	.44	.41	.49	.45	.41
Q4, child 6 months		.59	.52		.54	.47		.55	.49
Q5 child 18 months			.57			.52			.54
Q6 child 3 years									

**Table 5.** Rosenberg Self-Esteem Scale. Explained variance and correlation between the short form scores for each step in the linear regression analysis. One new item is added for each step.

Step #	Item added	% explained variance of total score	Correlation with total score	Alpha reliability
1	I take a positive attitude toward myself	64	.80	
2	I certainly feel useless at times	78	.87	.65
3	I feel I do not have much to be proud of	85	.92	.74
4	I feel that I am a person of worth, at least on an equal plane with others	90	.95	.80

Analyses of data material from 250 adolescents aged 17-18 years made available by Mette Ystgaard (1993). Alpha reliability for the full scale (10 items): 0.88.

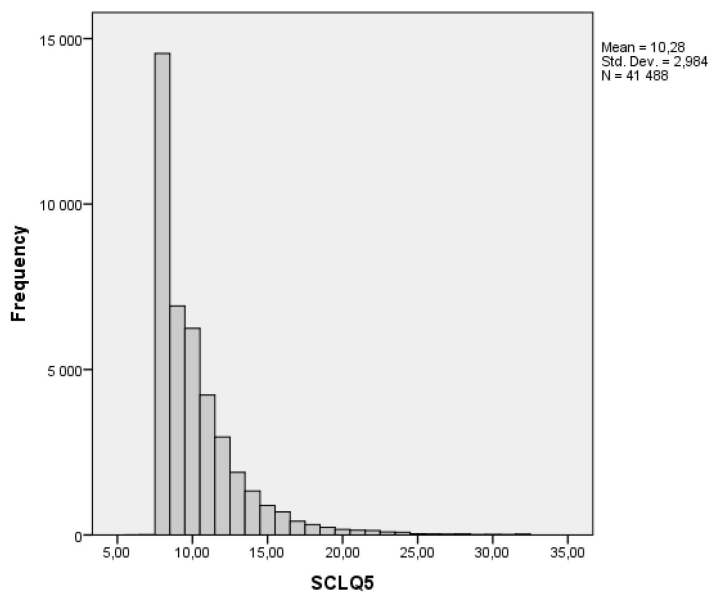
years. The correlation between the SCL-12 and the full SCL-25, based on the original data used to select the items, was 0.97. The correlation between the 8-item depression score from the SCL-12 and the full SCL depression score was also 0.97. The Alpha reliability was 0.90 for SCL-12 and 0.86 for the 8-item depression score. The anxiety scale in the SCL-12 is identical with the anxiety scale in the SCL-8.

The distribution of the SCL-10 scores is shown in Figure 1. The scores range from 10 (score 1, meaning no symptoms on any items) to 40 (scores 4, meaning "extremely bothered", on all items).

The list of items from the *Rosenberg Self-Esteem Scale (RSES)* is shown in Table 5. One single item, "I take a positive attitude toward myself", correlates at 0.80 with the full scale and explains 0.64% of the full scale variance. The sum of two items correlates 0.87 with the full scale. The correlation increases to 0.92 for a three item short-form version and to 0.95 for a four item version. Cronbach alpha for the four item version was 0.80. Alpha values observed with the MoBa data ranged from 0.74 (questionnaire week 15 of pregnancy) to 0.79 (questionnaire at 36 months).

Correlations between RSES scores in the various MoBa questionnaires are shown in Table 6. They vary from 0.57 for the longest time lag to around 0.65 for the shortest time lags. The distribution of the RSES scores, observed in MoBa questionnaire 5, is shown in Figure 2. The scores range from 4 (all score) to 16 (all scores 4).

The Generalized Self-Efficacy scale was subjected to the same analytic procedure as the RSES. As the GSE has been used less extensively in previous population based health studies, however, we performed some additional analyses of its underlying covariance

**Figure 1.** Distribution of SCL-8 scores from MoBa questionnaire 5.**Table 6.** Correlations between RSES scores at different points in time.

	Q3	Q4	Q5	Q6
Q1, pregnancy week 15	.66	.61	.59	.57
Q3, pregnancy week 30		.62	.60	.58
Q4, child 6 months			.64	.60
Q5 child 18 months				.65
Q6 child 3 years				

**Table 7.** Generalized Self-Efficacy scale (GSE). Explained variance and correlation between the short form scores for each step in the linear regression analysis. One new item is added for each step.

Step #	Item added	% explained variance of total score	Correlation with total score	Alpha reliability
1	I can remain calm when facing difficulties because I can rely on my coping abilities	56	.75	
2	I am confident that I could deal efficiently with unexpected events	74	.86	.66
3	I can always manage to solve difficult problems if I try hard enough	83	.91	.70
4	If I am in trouble, I can usually think of a solution	89	.94	.75
5	If someone opposes me, I can find the means and ways to get what I want	92	.96	.78

Alpha reliability for the full scale (10 items): 0.88

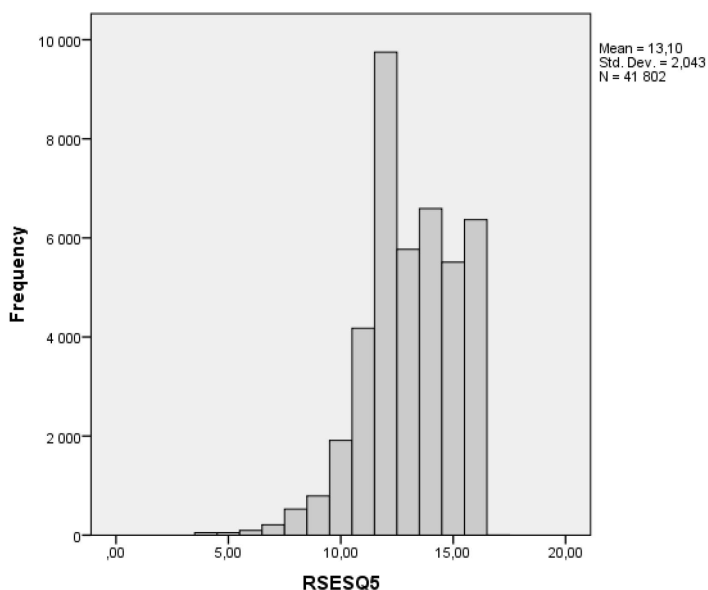
structure. A stepwise multiple regression analysis of the first five items resulted in multiple R=0.96, thus explaining 92% of the variance of the full scale index (see Table 7). A sum-score index of the five item short version (GSE-5) correlated 0.96 with the full version, and the Cronbach alpha of the short scale was 0.78. In MoBa, the alpha reliabilities for GSE-5 were 0.84 at both Q3 and Q5.

Next, an exploratory factor analysis supported the uni-dimensionality of the full scale, yielding Eigenvalues of 4.73 and 0.83 for the first two factors, respectively. To further examine the structural validity we conducted confirmatory factor analyses (CFA) of the full and short version scales, using the Mplus software. The hypothesized model with one latent factor showed acceptable fit for the full scale ( $\chi^2=379.27$ ,  $df=35$ , RMSEA=0.079, CFI=0.92) and even better fit for the short scale ( $\chi^2=46.26$ ,  $df=5$ , RMSEA=0.073, CFI=0.97). Thus, in terms of model fit the short scale showed high structural validity and was superior to the full scale.

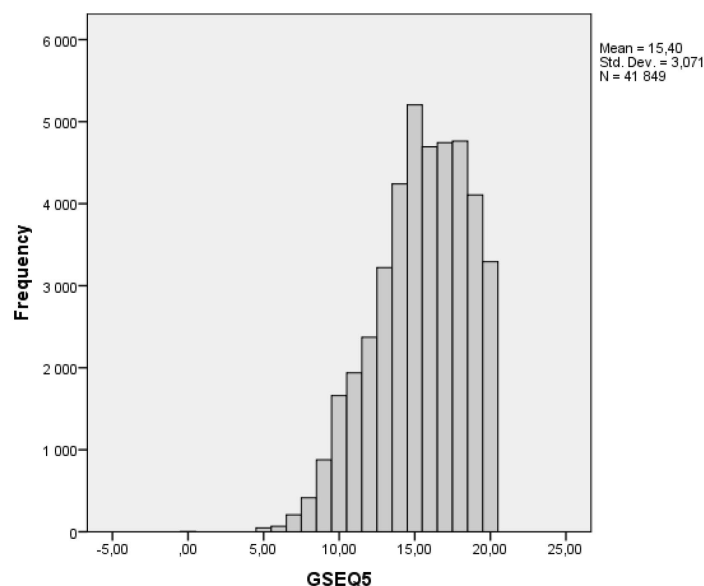
The distribution of the GSE scores does not dramatically deviate from a normal distribution even after reduced to five items, as shown in Figure 3 (skewness: -0.24; kurtosis: -0.55). The intercorrelation between the GSE scores over a time-lag around 19 months was 0.61.

**DISCUSSION**

This paper was meant to serve as psychometric documentation for researchers using these MoBa data and, more generally, to illustrate that the majority of psychometric instruments can be abbreviated without a loss in measurement precision which badly limits their usability. Many such instruments are used for clinical purposes or health screening. Here even relatively highly valid instruments will produce more false positives than true positives for rare disorders or characteristics, requiring that the reliability is maximized even if it costs a large number of items. For epidemiological studies of risk and protective factors, however, the reliability of most of the original instruments is more than good enough and can be somewhat reduced without much loss of information. In the examples of short-form scales described here, the



**Figure 2.** Distribution of RSES scores.



**Figure 3.** Distribution of GSE scores.

reliability is reduced with around 10% or less, which means very little for the risk estimates. Besides, these estimates can be adjusted for imperfect reliability. How far to go in abbreviating each instrument, depends on how they are intended to be used. If a variable is used as one of several predictors, we will usually tolerate a bigger loss in measurement precision than if the same variable is used as the principal predictor or as the outcome measure.

Even if the alpha reliability for our three examples of short-form instrument came out acceptably, there is a tendency for short-form instruments developed through stepwise linear regression as applied here, to show relatively low internal consistency reliability. The reason is that most phenomena assessed by psychometric instruments are by nature heterogeneous. Mental distress as measured in SCL-25 consists of symptoms of anxiety as well as depression, and even within each type of disorder the symptoms are not perfectly homogeneous. Having panic spells is for instance not highly correlated with headache, although both are included as anxiety symptoms in the SCL-25. The regression analyses “scans for diversity” in the sense that adding a new item to a set of short-form items already selected, usually explains more of the remaining variance in the total score if it measures something different, than if it measures approximately the same as the items already included. Suppose for instance we should want to measure mental distress (consisting of anxiety and depression) with only two items, and we would have to trust subjective judgment alone. It would then make sense to pick one item on depression and one on anxiety rather than two anxiety or two depression items, even if choosing one of each would probably reduce internal consistency. Comparing our observed alpha values with the rather low “test-retest” correlations would seem to indicate the opposite; that internal consistency is better than test-retest in our short-form measures. That is because our stability statistics are probably even worse underestimates of reliability than are the alpha measures. The “test-retest” values reflect the long time-lag between the observations and are influenced also by true changes in the phenomena under study.

Most of the short-form instruments should be used with some care. Scores from many psychometric instruments have a skewed distribution, and reducing the number of items tends to increase this skewedness. That implies that often such scores need to be transformed, for instance to logarithmic scores, to obtain distributions closer to a normal distribution. A skewed distribution also often implies a bad sensitivity in one end of the scale. For instance, as indicated by Figure 1, one third of the respondents did not report any symptoms at all, meaning that SCL-8 is not at all suitable to differentiate between good and very good mental health. But then again, the original SCL-25 is also not very suitable for this purpose, and it was never meant to be.

One may object that the selection of items for short-form instruments should apply more statistically sophis-

ticated methods, like item response theory (IRT). The short-form instruments chosen for the MoBa were developed before, or a few years after the start-up of the study in 1999, when IRT was not very well known. However, although IRT may have some advantages, we think the results show that not much can be gained by using more advanced methods.

There is an ongoing debate about the challenges of constructing short-form instruments, and about the optimal analytic strategy (10,11). Several studies have used factor analytic approaches (12,13) instead of the regression method applied here, and there are advantages and limitations to each strategy. Yet, if the main aim is to obtain a short scale with maximum correlation with the full scale the regression approach might be preferable (albeit with the possible cost of reduced internal reliability).

Although short scales typically are seen as sub-optimal compared to the full scales, it is noteworthy that in principle a short scale might represent an improvement to the full scale. If the full scale includes ambiguous, redundant or outdated items, a shorter scale that retains the best items might perform as well, or better, on some criteria. Our analyses of the GSE (see results) serves as an example in which the short version had somewhat reduced reliability, but was superior in terms of improved model fit and structural validity.

Copyright concerns may represent a practical obstacle for using abbreviated instruments. Often scales judged to be suitable for inclusion in epidemiological studies may be subject to commercial business. In our experience it is usually difficult to have permission to abbreviate scales in commercial use. But including copyright regulated scales in their original form is also usually not an option, simply because paying for using them becomes too expensive.

Traditionally, many editors and reviewers of journals within psychology, epidemiology and psychiatry have been skeptical to self-report scales, and in particular to self-made self-report scales. Experiences among researchers having used the MoBa short-form data are mixed. For the most part peer-reviewers have accepted the abbreviated measures in MoBa and in other population based studies like HUNT, although reduced alpha reliability sometimes have raised eyebrows. Our impression is that reviewers not very familiar with psychometrics are generally among the most critical. Our impression is also that the understanding of quantitative methodology and psychometric principals is improving among behavioral and psychiatric and medical researchers. We think such an increased insight makes it easier to realize that a measure which correlates 0.95 with a “gold standard” instrument cannot be completely useless.

#### ACKNOWLEDGEMENT

We are grateful to Mette Ystgaard and to the former National Health Screening Services, who made their data materials available, and to Inger Sandanger, who gave her permission to re-publish some previously published results.

**REFERENCES**

1. Hesbacher PT, Rickels R, Morris RJ, Newman H, Rosenfeld MD. Psychiatric illness in family practice. *J Clin Psychiatry* 1980; **41**: 6-10.
2. Derogatis LR. *SCL-90-R: Administration, scoring and procedures manual*. National Computer Systems, Minneapolis, 1994.
3. Rosenberg M. *Society and the Adolescent Self-Image*, revised edition. Middletown, CT: Wesleyan University Press, 1989.
4. Schwarzer R. *Self-efficacy: Thought control of action*. Washington, DC: Hemisphere, 1992.
5. Schwarzer R. Self-efficacy can moderate the planning-behaviour relationship. *Psychol Health* 2009; **24** (S1): 43.
6. Leganger A, Kraft P, Røysamb E. Perceived self-efficacy in health behaviour research: Conceptualisation, measurement and correlates. *Psychol Health* 2000; **15**: 51-69.
7. Tambs K, Moum T. How well can a few questionnaire items indicate mental health? *Acta Psychiatr Scand* 1993; **87**: 364-367.
8. Ystgaard M. *Sårbar ungdom og sosial støtte. En tilnærming til forebygging av psykisk stress og selvmord*. Oslo: Senter for sosialt nettverk og helse, 1993.
9. Røysamb E, Strype J. Optimism and pessimism: Underlying structure and dimensionality. *J Soc Clin Psychol* 2002; **21**: 1-19.
10. Emons WHM, Sijtsma K, Meijer RR. On the consistency of individual classification using short scales. *Psychol Methods* 2007; **12**: 105-20.
11. Smith GT, McCarthy DM, Anderson KG. On the sins of short-form development. *Psychol Assess* 2000; **12**: 102-11.
12. Marsh HW, Ellis LA, Parada RH, Richards G, Heubeck BG. A short version of the Self Description Questionnaire II: Operationalizing criteria for short-form evaluation with new applications of confirmatory factor analyses. *Psychol Assess* 2005; **17**: 81-102.
13. Donnellan MB, Oswald FL, Baird BM, Lucas RE. The Mini-IPIP scales: Tiny-yet-effective measures of the big five factors of personality. *Psychol Assess* 2006; **18**: 192-203.