The Prospective and Retrospective Memory Questionnaire (PRMQ): Normative data and latent structure in a large non-clinical sample

John R. Crawford and Geoff Smith
University of Aberdeen, UK

Elizabeth A. Maylor
University of Warwick, UK

Sergio Della Sala and Robert H. Logie
University of Aberdeen, UK

The Prospective and Retrospective Memory Questionnaire (PRMQ; Smith, Della Sala, Logie, & Maylor, 2000) was developed to provide a self-report measure of prospective and retrospective memory slips in everyday life. It consists of sixteen items, eight asking about prospective memory failures, and eight concerning retrospective failures. The PRMQ was administered to a sample of the general adult population (N = 551) ranging in age between 17 and 94. Ten competing models of the latent structure of the PRMQ were derived from theoretical and empirical sources and were tested using confirmatory factor analysis. The model with the best fit had a tripartite structure and consisted of a general memory factor (all items loaded on this factor) plus orthogonal specific factors of prospective and retrospective memory. The reliabilities (internal consistency) of the Total scale and the Prospective and Retrospective scales were acceptable: Cronbach's alpha was 0.89, 0.84, and 0.80, respectively. Age and gender did not influence PRMQ scores, thereby simplifying the presentation and interpretation of normative data. To ease interpretation of scores on the PRMQ, tables are presented for conversion of raw scores on the Total scale and Prospective and Retrospective scales to T scores (confidence limits on scores are also provided). In addition, tables are provided to allow users to assess the reliability and abnormality of differences between an individual's scores on the Prospective and Retrospective scales.

Remembering to do things at the appropriate time or in response to an appropriate cue event is as important in everyday life as is remembering information from the past. Prospective memory is concerned with the timing of when things are remembered, whereas retrospective memory is concerned with what should be remembered. From this it can be seen that prospective memory would fail if we remembered we were to do something at a given time, but not what that something was. Thus, prospective memory is distinct yet not entirely independent of retrospective memory. Failures of prospective memory can have various consequences. Missing an anticipated television programme may be annoying, but failing to take medication could be very serious. Because of the importance of certain failures, and their frequent occurrence in the healthy popula-

Requests for reprints should be sent to Professor John R. Crawford, Department of Psychology, King's College, University of Aberdeen AB24 2UB, UK. Email j.crawford@abdn.ac.uk

This research was supported by The Wellcome Trust (grant 047065) and the Medical Research Council (grant G9606610N). Versions of the questionnaire are available in different languages thanks to European Commission contract number SOC 97 201791 05F03.

http://www.tandf.co.uk/journals/pp/09658211.html DOI:10.1080/09658210244000027

tion (Dobbs & Rule, 1987), there is growing interest in the systematic study of prospective memory; see Brandimonte, Einstein, and McDaniel (1996), or the recent special issue of *Applied Cognitive Psychology* (Ellis & Kvavilashvili, 2000) for recent discussion.

One way to gain insight into any distinction between prospective and retrospective memory performance in everyday life is through the use of self-rating questionnaires. This may allow systematic study of whether failures are prospective or retrospective in nature (Smith et al., 2000) and investigation of whether different variables are associated with prospective versus retrospective failures. Smith et al. (2000) note that previous work using self-reports of memory ability have all but ignored any distinction between prospective and retrospective memory. For example the Cognitive Failures Questionnaire (Broadbent, Cooper, Fitzgerald, & Parkes, 1982) includes only 2 items out of 25 that can be said to ask about prospective memory. Also the Everyday Memory Questionnaire developed by Sunderland, Harris, and Baddeley (1984; 1988) contains only 3 items out of 28 that probe prospective memory abilities. Smith et al. describe the Prospective and Retrospective Memory Questionnaire (PRMQ). This is a brief (16-item), self-report, measure of prospective and retrospective failures in everyday life. Eight of the items enquire about prospective memory and eight about retrospective memory. The items were also designed to contain an equal number concerned with either self-cued memory or environmentally cued memory, and with short-term versus long-term memory. Thus, although the primary aim was to develop a self-report scale that would systematically measure prospective and retrospective memory, each item can be categorised along three dimensions. For example, item 14 (‘‘If you tried to contact a friend or relative who was out, would you forget to try later?’’) is categorised as measuring prospective, long-term, self-cued memory; see Appendix 1 for a full list of items and their categorisation.

The present study had four principal aims. The first aim was to determine whether, in the general adult population, age and gender influence PRMQ scores. The second aim was to provide normative data for the PRMQ, including normative data on the magnitude of *discrepancies* between self-rated prospective and retrospective memory (i.e., the latter data will allow an assessment of the rarity or abnormality of the discrepancy between an individual's scores on the two scales). The third aim was to obtain estimates of the reliability of the PRMQ. Establishing the reliability of an instrument is a fundamental step in evaluating its potential utility. However, the reliability data can also be used (a) to provide confidence limits on individuals' scores, and (b) to test for reliable differences between individuals' scores on the Prospective and Retrospective scales. Finally, if the use of the PRMQ in research and practice is to be optimal, then it is necessary to have knowledge of the underlying structure of the instrument. The fourth aim of the present study was to evaluate competing models of the latent structure of the PRMQ, using confirmatory factor analysis (CFA). The parameterisation of CFA models operationalise hypotheses about the structure of the instrument. Therefore, to avoid undue repetition, the theoretical and methodological considerations that guided selection of these models are covered in the Method section.

METHOD

Participants

Complete PRMQ data were collected from 551 members of the general adult population (females = 344, males = 207). Participants were recruited from a wide variety of sources in and around two UK cities; these included commercial and public service organisations, community centres, and recreational clubs. There was a mix of urban and rural residents but, reflecting the distribution in the general population, the participants were predominantly urban dwellers. The mean age of the sample was 63.62 (SD = 15.59) with a range of 17 to 94 years. The mean years of education was 13.22 (SD = 3.38) with a range from 4 to 20 years.

Materials and procedure

Each potential participant received an introductory letter, a PRMQ form (see Appendix 1) and a form for recording demographic variables. The PRMQ was described as a set of questions about minor memory mistakes that everyone makes from time to time. Participants were asked to say how often each of these things happened to them on a 5-point scale: Very Often, Quite Often, Sometimes, Rarely, Never. Ratings were subsequently assigned numerical values of 5 (Very Often) to 1 (Never), resulting in minimum and maximum possible total scores of 16 and 80, respectively.

Confirmatory factor analysis

Confirmatory factor analysis (maximum likelihood) was performed on the variance–covariance matrix of the PRMQ items using EQS for Windows Version 5 (Bentler, 1995). The fit of CFA models was assessed using chi square (χ^2), the average off-diagonal standardised residual (AODSR), the root mean squared error of approximation (RMSEA; see Steiger, 2000) and the standardised root mean squared residual (SRMR; Bentler, 1995), and the Comparative Fit Index (CFI). Off-diagonal standardised residuals reflect the extent to which covariance between manifest variables has not been accounted for by the models under consideration. Values for the CFI can range from zero to unity; this index expresses the fit of a model relative to what is termed the null model (the null model posits no relationship between any of the manifest variables). Models in which the CFI is 0.95 or above are regarded as having good fit (Bentler & Mooijart, 1989). The RMSEA is widely recommended as a measure of fit because it penalises models that are not parsimonious and is sensitive to misspecified factor loadings (Hu & Bentler, 1998); an RMSEA <0.06 is taken as indicating good fit (Hu & Bentler, 1999). The SRMR is a relatively recently developed index and is particularly sensitive to misspecified factor covariances; an SRMR of < 0.08 is regarded as indicating good fit (Hu & Bentler, 1999).

A model is considered to be nested within another model if it differs only in imposing additional constraints on the relationships between variables specified in the initial model. The difference between chi square for nested models is itself distributed as chi square with k degrees of freedom where k equals the degrees of freedom for the more constrained model minus the degrees of freedom for the less constrained model. Because of this it is possible to test directly whether more constrained models have significantly poorer fits than the less constrained models; this feature of confirmatory factor analysis (CFA) is one of its major advantages over exploratory factor analysis (EFA).

The first model (Model 1) to be evaluated was a single factor model; this model expressed the hypothesis that the variance in the PRMQ can be partitioned into one general factor plus error variance associated with each individual item (error variance here refers to the combination of true variance in the item that is independent of the factor, plus random error). It is standard practice to test the fit of a one-factor model because it is the most parsimonious of all possible models.

Model 2a expressed the hypothesis that the PRMQ measures two independent (i.e., orthogonal) factors, prospective and retrospective memory. Therefore, in this model all prospective and retrospective items were constrained to load only on their respective factors. As noted, the PRMQ items can also be categorised as to whether they are concerned with short- or long-term memory, and self-cued memory or environmentally cued memory. As these distinctions may be an important source of covariance among PRMQ items, independent factor models representing them were also constructed. Thus, Model 2b consisted of two independent factors made up of short-term and long-term memory items respectively. Model 2c consisted of two independent self-cued memory and environmentally cued memory factors.

Models 3a to 3c were parameterised exactly as their Model 2 counterparts except that the factors in each were allowed to correlate; e.g., Model 3a tests the hypothesis that two factors, prospective and retrospective memory, explain the covariance among items but that these factors are not independent (i.e., orthogonal).

Models 4a to 4c represented variants on the hypothesis that the PRMQ has a tripartite structure. The basic tripartite model (Model 4a) was parameterised so that all 16 items were indicators of a common factor (representing general self-rated memory). In addition, the eight prospective items were also indicators of a factor reflecting the variance specific to prospective memory and the eight retrospective items were indicators of a specific retrospective memory factor. The specific factors were constrained to be orthogonal to each other and to the common factor. Two more constrained variants on this model were also constructed. In Model 4b the retrospective factor was omitted entirely; this model therefore posits that all PRMQ items tap general memory, but only the prospective items tap an additional specific factor. Model 4c was the reverse of Model 4b, consisting of a general memory factor and specific retrospective factor (i.e., the prospective factor was omitted). It will be appreciated that, if both specific factors were omitted, this would recreate the single factor model (Model 1).

Normative data

In presenting normative data for the PRMQ, it was considered desirable that scores on the Prospective and Retrospective scales and Total scale should be expressed on a common metric. Therefore tables were constructed to convert raw scores on the PRMQ to *T* scores (*T* scores have a mean of 50 and a *SD* of 10). *T* scores were chosen because they are in widespread use, their meaning is easy to convey, and they permit users of the scale to rapidly assimilate an individual's or group's standing (Crawford, Venneri, & O'Carroll, 1998b). For example, if an individual obtained *T* scores of 40 and 35 on the Prospective and Retrospective scales respectively, it can immediately be seen that he or she is 1 *SD* below the estimated population mean on the Prospective scale and 1.5 *SD*s below the mean on the Retrospective scale.

When working with individuals' scores it is important that test users are aware that all psychological test scores are fallible. For this reason, and in order to quantify the extent of this fallibility, it is widely recommended that scores should be accompanied by confidence limits (Nunnally & Bernstein, 1994). Confidence limits on scores for each of the three PRMQ scales were formed by obtaining the standard error of measurement for true scores (Glutting, Mcdermott, & Stanley, 1987; Stanley, 1971) using the following formula:

$$SEM_{xt} = r_{xx}\left(S_x\sqrt{1-r_{xx}}\right), \qquad (1)$$

where S_x is the standard deviation of the scale (10 in the present case as raw scores are converted to *T* scores) and r_{xx} is the reliability of the scale (normally estimated using Cronbach's alpha). Confidence limits are formed by multiplying the SEM_{xt} by a value of z (a standard normal deviate) corresponding to the desired confidence limits; for 95% limits (the most commonly used) the SEM is multiplied by 1.96. These confidence limits are not symmetrical around individuals' *obtained* scores but around their estimated *true* scores (Nunnally & Bernstein, 1994; Silverstein, 1989; Stanley, 1971). The estimated true score is obtained by multiplying the obtained score, in deviation form, by the reliability of the test. It can be seen then that estimated true scores are regressed towards the mean, the extent of this regression varying inversely with the reliability of the scale. The formula is as follows:

$$\text{True score} = r_{xx}(X - \overline{X}) + \overline{X}, \qquad (2)$$

where X is the obtained score and $\overline{X}$ is the mean for the scale. Thus, for example, if an individual obtained a score of 40 on a scale that had a mean of 50 and a reliability of 0.8, the individual's estimated true score would be 42.

In addition to standard normative data, it would be useful for users to have some means of evaluating *discrepancies* between an individual's Prospective and Retrospective scores. One approach to this is to provide a method of testing whether scores are *reliably* different from each other. Stanley (1971) and others (e.g., Silverstein, 1989) recommend that this is done using estimated true scores rather than obtained scores. Critical values for the difference between an individual's estimated true scores are obtained by first calculating the standard error of the difference:

$$SE_{D_t} = \sqrt{SEM^2_{x_t} + SEM^2_{y_t}}, \qquad (3)$$

where SEM_{x_t} and SEM_{y_t} are the standard errors of measurement for true scores obtained using formula (2). Critical values are obtained by multiplying the standard error of the difference for true scores (SE_{D_t}) by the value of z (a standard normal deviate) corresponding to the required significance level (i.e., 1.96 for the 0.05 level).

In many circumstances (e.g., in potential clinical use of the PRMQ) it would also be useful to have information on the *rarity* or *abnormality* of the observed discrepancy; the distinction between the reliability and the abnormality of a difference is elaborated on in the Discussion section. To provide this we used a method developed by Crawford, Howell and Garthwaite (1998a). This method uses the following formula to obtain a quantity that is distributed as t:

$$t = \frac{|T_x - T_y|}{\sqrt{\left[S_x + S_y - S_xS_yr_{xy}\right]\left(\dfrac{N+1}{N}\right)}}, \qquad (4)$$

where T_x and T_y are the individual's *T* scores on the two scales being compared, S_x and S_y are the standard deviations of the scales (10 in the present case as *T* scores are used), r_{xy} is the correlation between the scales, and N is the size of the normative sample. The percentile point corresponding to the t obtained from this formula is then found and multiplied by 100 to provide an estimate of the percentage of the population equalling or exceeding the observed discrepancy. (To obtain the percentage equalling or exceeding the observed discrepancy, *regardless of the sign of the* *discrepancy*, the percentile point is multiplied by two before being multiplied by 100.)

RESULTS

Reliability of the PRMQ

The reliability (internal consistency) of the PRMQ was estimated using Cronbach's alpha. Cronbach's alpha was 0.89 (95% CI = .88 to .90) for the Total scale, 0.84 (95% CI = .82 to .86) for the Prospective scale, and 0.80 (95% CI = .77 to .82) for the Retrospective scale.

Influence of age and gender on PRMQ scores

Independent samples *t*-tests revealed that the mean scores of females and males did not differ significantly on the Total scale ($t = 1.23, p = .22$) or Prospective scale ($t = 0.47, p = 0.64$). A significant effect of gender was observed for the Retrospective subscale ($t = 2.73, p < .01$); females reported fewer memory lapses than males. However, the effect size was very modest with gender accounting for only 1.3% of the variance; therefore the significance level reflects the high statistical power conveyed by the large sample size. The Pearson correlations between age and scores on the Total scale ($r = .08$), Prospective scale ($r = .08$), and Retrospective scale ($r = .07$) did not achieve statistical significance. As, for all practical purposes, these demographic variables did not influence scores, presentation of normative data is simplified as there is no need for stratification by these variables.

Summary statistics and normative data for the PRMQ

The means, medians, *SD*s and ranges for the Total scale and the Prospective and Retrospective subscales are presented in Table 1. The fourth column of Table 1 presents the standard errors of measurement of true scores on the scales when scores are expressed as *T* scores (see later section). Table 2 permits conversion of raw scores on the Total scale to *T* scores. Estimated true scores and the 95% lower and upper confidence limits on true scores are also presented. Tables 3 and 4 present the equivalent information for the Prospective and Retrospective scales, respectively. Table 5 can be used to determine whether the difference between an individual's *true* scores on the Prospective and Retrospective scales are reliably different; it presents critical values for four levels of significance (.15, .10, .05 and .01). The values in this table were obtained using formula (3).

TABLE 1
Summary statistics for the PRMQ

Scale	*Mean*	*SD*	*Range*	SEM_{x_t}
Total PRMQ score	38.88	9.15	17–67	2.95
Prospective score	20.18	4.91	8–35	3.36
Retrospective score	18.69	4.98	8–33	3.58

Table 6 presents data on the *abnormality* of differences between the Prospective and Retrospective scales. It records the percentage of the population expected to equal or exceed a given difference between *T* scores (note *T* scores are used with this table *not* true scores). The values in this table were generated using formula (4). Table 6 has two columns: the first column records the percentage expected to equal or exceed a given difference *in a specified direction*; the second column records the percentage expected *regardless of the sign of the difference*.

Testing competing confirmatory factor analytic models of the PRMQ

The fit statistics for all of the CFA models are presented in Table 7. It can be seen that the general factor model (Model 1) had poor fit; the model's chi square is large and the CFI is low. However, the item loadings on this factor were high; the averaged loading was 0.59. Thus, although the hypothesis that the PRMQ is a unidimensional measure is untenable, there is nevertheless evidence for substantial common variance among the items.

Models 2a to 2c expressed the hypotheses that the PRMQ measures two independent factors of prospective and retrospective memory (Model 2a), short-term and long-term memory (Model 2b), or self-cued memory and environmentally cued memory (Model 2c). The fit of all these models was very poor and considerably worse than that of the single factor model; chi square is large and the CFI is low in all three cases.

The fits of the correlated factors models (Models 3a to 3c) are markedly superior to their

TABLE 2
Table for converting raw scores on the PRMQ Total scale to *T* scores and for obtaining 95% confidence limits on true scores

			95% confidence limits					*95% confidence limits*	
Raw score	*T Score*	*True score*	*Lower*	*Upper*	**Raw score**	*T Score*	*True score*	*Lower*	*Upper*
16	74	72	66	78	**49**	39	40	34	46
17	74	71	65	77	**50**	38	39	33	45
18	73	70	64	76	**51**	37	38	32	44
19	72	69	63	75	**52**	36	37	31	43
20	71	68	63	74	**53**	35	36	30	42
21	70	67	62	73	**54**	34	35	30	41
22	69	67	61	72	**55**	33	34	29	40
23	68	66	60	71	**56**	32	34	28	39
24	67	65	59	71	**57**	31	33	27	38
25	66	64	58	70	**58**	30	32	26	38
26	65	63	57	69	**59**	28	30	24	36
27	63	61	55	67	**60**	27	29	23	35
28	62	60	55	66	**61**	26	28	22	34
29	61	59	54	65	**62**	25	27	22	33
30	60	59	53	64	**63**	24	26	21	32
31	59	58	52	63	**64**	23	26	20	31
32	58	57	51	63	**65**	22	25	19	30
33	57	56	50	62	**66**	21	24	18	30
34	56	55	49	61	**67**	20	23	17	29
35	55	54	48	60	**68**	19	22	16	28
36	54	53	47	59	**69**	18	21	15	27
37	52	51	46	57	**70**	16	19	14	25
38	51	51	45	56	**71**	15	18	13	24
39	50	50	44	55	**72**	14	18	12	23
40	49	49	43	55	**73**	13	17	11	22
41	48	48	42	54	**74**	12	16	10	22
42	47	47	41	53	**75**	11	15	9	21
43	46	46	40	52	**76**	10	14	8	20
44	45	45	39	51	**77**	9	13	7	19
45	44	44	38	50	**78**	8	12	6	18
46	43	43	38	49	**79**	7	11	6	17
47	42	43	37	48	**80**	5	10	4	15
48	40	41	35	47					

T scores were calculated from reflected raw scores so that low scores reflect poor self-rated memory.

independent factors counterparts. As noted, inferential statistics can be applied to compare nested models. Models 2a to 2c are nested within Models 3a to 3c, respectively, in that they differ only by the imposition of the constraint that the factors are independent. The results from chi square difference tests used to compare these nested models are presented in Table 8. It can be seen that the correlated factors models had significantly better fit ($p < .001$) than their independent factors counterparts; this demonstrates that the hypothesis of independence between the scales is untenable. However, it can be seen from Table 7 that the fit statistics for Models 3b and 3c are very similar (indeed, identical for most indices) to those of the more parsimonious one-factor model. Although at first glance it may appear that the one-factor model is very different from the correlated factors models, it is nested within them; Models 3a to 3c can be rendered equivalent to Model 1 by constraining the correlations between the factors to unity (i.e., $r = 1.0$). Comparison of Model 1 with Models 3b and 3c revealed that the fit of these latter models was not significantly better than Model (Table 8). In contrast, comparison of Model 1 with Model 3a yielded a highly significant difference in fit ($p < .001$). This latter result provides further evidence that it is also untenable to view the PRMQ as measuring only a unitary memory factor.

The fit of the correlated prospective and retrospective factors model was markedly superior to the other competing models considered so far. However, as can be seen from Table 7, this model

TABLE 3
Table for converting raw scores on the PRMQ Prospective scale to *T* scores and for obtaining 95% confidence limits on true scores

			95% Confidence limits on true score	
Raw score	*T score*	*True score*	*Lower*	*Upper*
8	74	70	64	77
9	72	69	62	75
10	70	67	61	74
11	68	65	59	72
12	66	64	57	70
13	64	62	55	69
14	62	60	54	67
15	60	59	52	65
16	58	57	50	64
17	56	55	49	62
18	54	54	47	60
19	52	52	45	59
20	50	50	44	57
21	48	49	42	55
22	46	47	40	54
23	44	45	39	52
24	42	44	37	50
25	40	42	35	48
26	38	40	34	47
27	36	39	32	45
28	34	37	30	43
29	32	35	29	42
30	30	33	27	40
31	28	32	25	38
32	26	30	24	37
33	24	28	22	35
34	22	27	20	33
35	19	24	18	31
36	17	23	16	29
37	15	21	14	27
38	13	19	13	26
39	11	18	11	24
40	9	16	9	22

TABLE 4
Table for converting raw scores on the PRMQ Prospective scale to *T* scores and for obtaining 95% confidence limits on true scores

			95% Confidence limits on true score	
Raw score	*T score*	*True score*	*Lower*	*Upper*
8	71	67	60	74
9	69	65	58	72
10	67	64	57	71
11	65	62	55	69
12	63	61	54	68
13	61	59	52	66
14	59	57	50	64
15	57	56	49	63
16	55	54	47	61
17	53	53	46	60
18	51	51	44	58
19	49	49	42	56
20	47	48	41	55
21	45	46	39	53
22	43	45	38	52
23	41	43	36	50
24	39	41	34	48
25	37	40	33	47
26	35	38	31	45
27	33	37	30	44
28	31	35	28	42
29	29	33	26	40
30	27	32	25	39
31	25	30	23	37
32	23	29	22	36
33	21	27	20	34
34	19	25	18	32
35	17	24	17	31
36	15	22	15	29
37	13	21	14	28
38	11	19	12	26
39	9	17	10	24
40	7	16	9	23

still includes misspecifications; chi square is large and, despite exceeding the recommended minimal criteria for a practical level of fit (0.9), the CFI is still relatively modest.

The basic tripartite model (Model 4a) had by far the best fit of all the models reported so far. The CFI of 0.95 is high and the AODSR is low (0.023). A model with a CFI greater than or equal to 0.95 and an SMR equal to or smaller than 0.08 (or RMSEA equal to or smaller than 0.06) indicates good fit (Hu & Bentler, 1999); it can be seen that the tripartite model fulfils this criterion. Further, although statistically significant, chi square for this model (245.4) is substantially less than that for all other models. A

TABLE 5
Critical values for significant (i.e., reliable) differences between estimated true scores on the Prospective and Retrospective scales

	Significance level			
	.15	.10	.05	.01
Two-tailed critical value	7	8	10	13
One-tailed critical value	5	6	8	11

Estimated true scores for the Prospective and Retrospective scales should be obtained from columns 3 of Tables 3 and 4, respectively.

TABLE 6
Percentage of the population estimated to exhibit discrepancies as or more extreme than a given discrepancy between *T* scores on the Prospective and Retrospective scales

	Percentage as or more extreme	
Discrepancy (ignoring sign)	*Directional difference*	*Absolute difference*
1	44.8	89.5
2	39.6	79.2
3	34.6	69.3
4	29.9	59.8
5	25.5	51.0
6	21.5	42.9
7	17.8	35.6
8	14.6	29.2
9	11.8	23.6
10	9.4	18.8
11	7.4	14.7
12	5.7	11.4
13	4.4	8.7
14	3.3	6.5
15	2.4	4.8
16	1.8	3.5
17	1.3	2.5
18	0.9	1.8
19	0.6	1.3
20	0.4	0.9
21	0.3	0.6
22	0.2	0.4
23	0.1	0.3
24	0.1	0.2
25	0.1	0.1
26	0	0.1
27	0	0

T scores should be used with this table, *not* estimated true scores.

schematic representation of the standardised solution for Model 4a is presented as Figure 1. By convention, latent factors are represented by large ovals or circles and manifest (i.e., observed) variables as rectangles or squares. The manifest variables (i.e., PRMQ items) are identified by their item number and a mnemonic for the content of the item (see Appendix 1 for full details of items). Error variances are often represented by smaller ovals or circles (as they are also latent variables). However, the ovals are omitted here to simplify the appearance of the diagram; the causal arrows representing the influence of these variables on the manifest variables are retained. As can be seen from Figure 1, in the tripartite model all items are indicators of a common factor (general memory) but, in addition, the prospective and retrospective items are also indicators of specific prospective and retrospective memory factors.

Models 4b and 4c are nested within Model 4a and thus chi square difference tests can be used to determine if the additional constraints imposed by these models significantly worsen the fit. The results are presented in Table 8. The difference in chi square (60.9) when the specific retrospective memory factor was omitted (Model 4b) was highly significant. Similarly, when the specific prospective factor was omitted (Model 4c) the difference (43.4) was also highly significant. Therefore, both these factors, which represent variance specific to prospective and retrospective memory, are necessary for an adequate level of practical fit.

TABLE 7
Fit indices for confirmatory factor analytic models of the PRMQ (best fitting model in bold)

Model	χ^2	*df*	*AODSR*	*RMSEA*	*SRMR*	*CFI*
1. Single memory factor	407.2	104	0.043	0.073	0.057	0.89
2a. Prospective and retrospective memory as orthogonal factors	732.4	104	0.185	0.105	0.268	0.78
2b. Short- and long-term memory as orthogonal factors	971.6	104	0.197	0.123	0.288	0.70
2c. Self-cued and environmentally cued memory as orthogonal factors	979.1	104	0.199	0.124	0.290	0.69
3a. Prospective and retrospective memory as correlated factors	336.1	103	0.039	0.064	0.053	0.92
3b. Short- and long-term memory as correlated factors	406.5	103	0.043	0.073	0.057	0.89
3c. Self-cued and environmentally cued memory as correlated factors	406.6	103	0.043	0.073	0.057	0.89
4a. Tripartite model (general memory plus orthogonal specific prospective and retrospective factors)	**245.4**	**88**	**0.023**	**0.057**	**0.044**	**0.95**
4b. Tripartite model with specific retrospective factor removed	306.3	96	0.036	0.060	0.047	0.93
4c. Tripartite model with specific prospective factor removed	288.8	96	0.034	0.063	0.050	0.93

TABLE 8
Results of Testing for Differences Between Nested CFA Models of the PRMQ

Comparison				
More constrained	*Less constrained*	*Difference in* χ^2	*df*	*p*
Model 2a	Model 3a	396.3	1	<.001
Model 2b	Model 3b	565.1	1	<.001
Model 2c	Model 3c	573.1	1	<.001
Model 1	Model 3a	71.1	1	<.001
Model 1	Model 3b	0.70	1	.40
Model 1	Model 3c	0.60	1	.44
Model 4b	Model 4a	60.9	8	<.001
Model 4c	Model 4a	43.4	8	<.001

DISCUSSION

Influence of demographic variables

One basic aim of the present study was to examine the influence of demographic variables on PRMQ scores. The modest effects of demographic variables further simplifies the use of the PRMQ as these need not be factored in when interpreting scores. The results also indicate that there was nothing to be gained by stratifying the sample by demographic variables when providing normative data.

The absence of significant correlations between age and PRMQ scores is obviously in sharp contrast with age-related decline on objective laboratory tests of both prospective and retrospective memory (see reviews by Maylor, 1996, and Zacks, 2000 respectively). Such age invariance in subjective memory ratings has been observed in many previous studies and has been attributed to several factors, notably the tendency to rate memory relative to one's peers (see Rabbitt, Maylor, McInnes, Bent, & Moore, 1995, for a summary of studies and detailed discussion). The present results would suggest that such factors apply to both prospective and retrospective items and therefore do not invalidate the main comparisons of interest as revealed by the modelling.

Competing models of the structure of the PRMQ

The use of CFA to test competing models of the latent structure of the PRMQ yielded clear-cut results. From the fit statistics presented in Table 7, it is clear that the hypothesis that the PRMQ measures a single factor (Model 1) is untenable, as is the conception of the PRMQ as a measure of two independent factors of prospective and retrospective memory (Model 2a). Although the correlated prospective and retrospective models had superior fit to the competing models discussed so far, they nevertheless contain substantial misspecifications as evidenced by the large chi square values and relatively modest CFI. In contrast, the full tripartite model (Models 4a) had a chi square value that, although statistically significant,[1] was relatively small; the correlated factors model implies the presence of a general factor whereas the tripartite model makes this general factor explicit. The tripartite model also allows for the fact that individual items vary in the extent to which they are indicators of the general versus specific factors.

More parsimonious versions of the tripartite models were tested to determine whether both of the specific factors were necessary to model the PRMQ adequately. These models posited that all items indexed a common memory factor but that only the prospective items (Model 4b) or retrospective items (Model 4c) indexed an additional specific factor. Had the fit of the tripartite model not deteriorated significantly when either of these factors were omitted it would have indicated either that the theory that inspired the model was flawed, or that the PRMQ items were inadequate indicators of the constructs, or both. However, chi square difference tests revealed that both specific components were necessary; the deterioration in fit when either was removed was substantial (see Table 8).

It would have been possible to specify and test more complex models of the structure of the PRMQ. Most obviously, models that *simultaneously* posited the existence of specific prospective /retrospective, self-cued/environmentally cued, and short-term/long-term factors could have been specified. However, evaluation of the models that specified the latter two of these sets of factors (Models 3b and 3c) did not provide a significantly better fit to the data than a single factor model. Therefore, the distinctions these factors represented were not a significant source of covariance among the PRMQ items and there was nothing to be gained in testing more complex models that incorporated them.

[1] When dealing with large sample sizes and a moderate number of items it is unusual to obtain non-significant chi square values for CFA models of self-report data (Byrne, 1994).

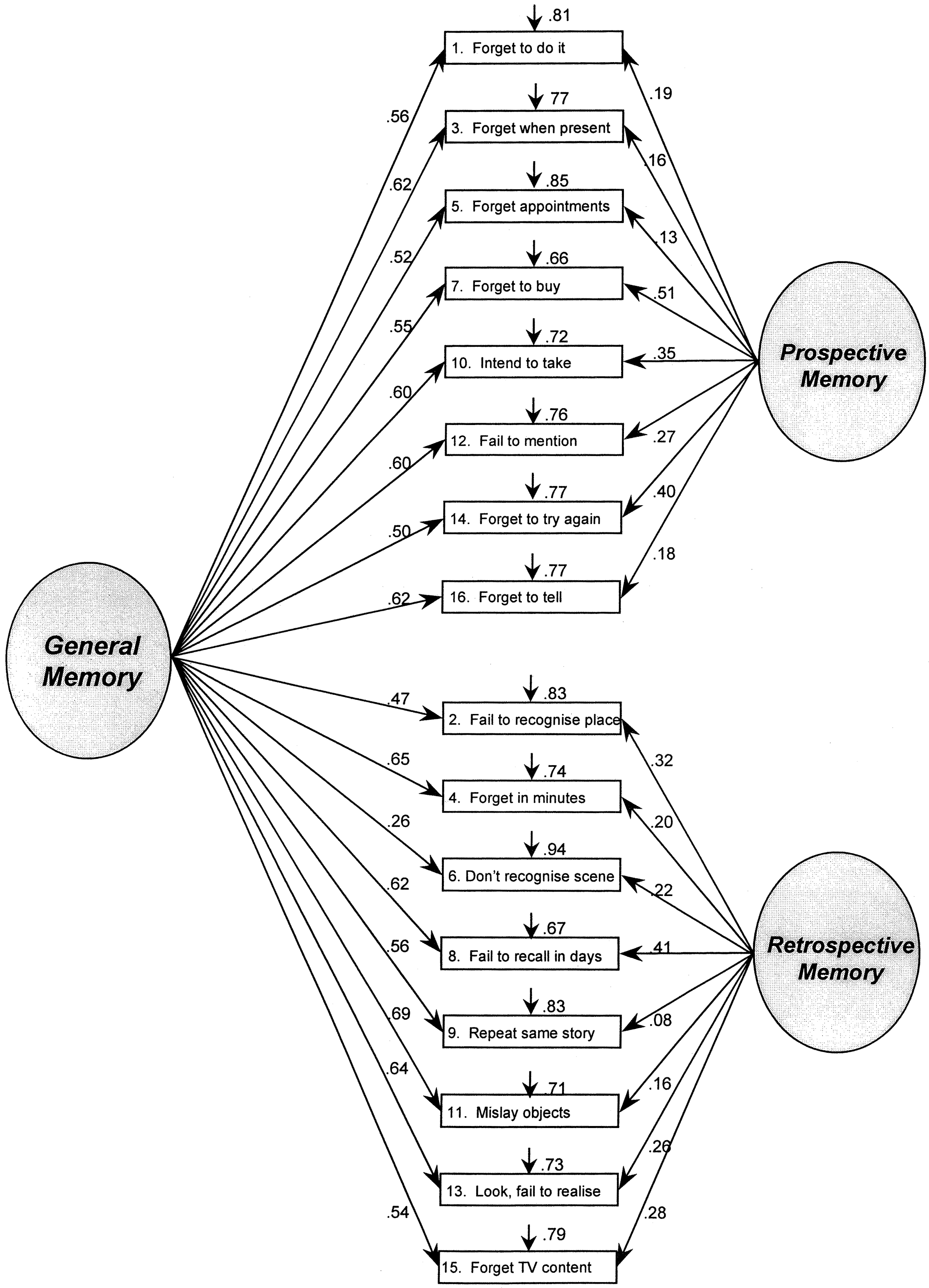

Figure 1. Graphical representation of the tripartite model of the PRMQ (Model 4a).

The failure to find an effect of self-cued versus environmentally cued items is of particular interest in the light of Craik's (1986) argument for the importance of this dimension in remembering. On the assumption that retrieval processes are more prone to failure in the absence of an external cue, Craik suggested that differences between prospective and retrospective remembering might arise because they are confounded respectively with self-generated and externally cued retrieval. Craik's arguments are primarily directed at explaining differences in performance on prospective and retrospective memory tasks. However, the present study offered an opportunity to determine whether the self-cued versus environmentally cued distinction exerts an impact on *self-rated* memory and, specifically, whether it can account for differences in self-rated prospective and retrospective memory.

In Smith et al.'s (2000) report on the PRMQ, memory lapses in situations requiring self-cueing were rated as significantly more frequent than in situations involving environmental cues. This result is consistent with Craik's (1986) functional account of the demands of different memory tasks. However, the present analysis of the latent structure of the PRMQ suggests that this effect reflects a quantitative rather than qualitative difference. That is, the effect of the self-cued versus environmentally cued distinction was not salient to participants when they rated their memory beyond producing a scaling (i.e., perceived difficulty) effect. Had this distinction exerted any additional explicit effect (or indeed implicit effect) then this would have left its signature in the pattern of covariances between items. However, factors that represented self-cued versus environmentally cued memory did not account for any more of the covariance between items than was explained by a single factor model (the fit statistics are essentially identical).

An analogy may help to clarify this distinction. Suppose that we collected ratings of the difficulty of simple arithmetic problems. Problems requiring operations that involved negative numbers, for example, would be likely to be rated as more difficult than problems that did not require these operations. However, if this difficulty effect was universal (i.e., if the perceived relative difficulty of these operations was equivalent across all individuals) then modelling of the covariance matrix of ratings would not uncover latent variables representing these operations. However, if there were individual differences in the perceived relative difficulty of tasks involving these operations (e.g., some individuals reported relatively less or more difficulty with negative numbers than others), then latent variables representing these operations would be necessary to model the data adequately.

To summarise, analysis of the PRMQ suggests that Craik's (1986) distinction between self-cued and environmentally cued memory also applies to self-rated memory in that items tapping self-generated recollection are rated as provoking more failures (Smith et al., 2000). However, the present CFA analysis suggests that these distinctions do not explain the pattern of covariances among items. This, of course, leaves open the possibility of further experimental exploration of how memory performance, rather than self-rated memory, is affected by the factors manipulated in the PRMQ; e.g., see Maylor, Darby, Logie, Della Sala, and Smith (in press), and Maylor, Smith, Della Sala, and Logie (2002).

As noted, the conclusion from the CFA modelling is that all the PRMQ items index a substantial common memory factor and that, in addition, they also index orthogonal dimensions that are specific to either prospective or retrospective memory. Therefore, it would be valid to use the PRMQ Total scale as a measure of general self-rated memory. In addition, it would also be valid to use the subscales as measures of self-rated Prospective and Retrospective memory, with the caveat that the researcher or clinician should be aware that the scales are a blend of variance *common* to both prospective and retrospective memory and variance *specific* to these constructs. When used with the individual case, the reliability and abnormality (i.e., rarity) of differences between the Prospective and Retrospective scales can be assessed using Tables 5 and 6 (see later section for details).

Encouraging validation of the PRMQ as a useful investigative tool comes from a recent study by Mantyla (2001) in which a group of middle-aged adults who specifically complained about everyday prospective memory problems were compared with a control group of non-complainers. On a Swedish translation of the PRMQ, there was indeed a significant interaction between group and prospective vs retrospective memory such that the group difference was much larger for the Prospective scale than for the Retrospective scale. Moreover, on a large battery of laboratory and naturalistic memory tasks, complainers were significantly worse on all the prospective memory tasks than the non-complainers, whereas the groups did not differ on any of the retrospective memory tasks. Thus, there is empirical evidence to suggest that relative differences across these two scales of the PRMQ may translate into actual performance differences in both laboratory and naturalistic settings.

Reliabilities and normative data

The reliabilities of the PRMQ scales, as measured by Cronbach's alpha, were .89, .84, and .80 for the Total scale and Prospective and Retrospective scales, respectively. The narrowness of the confidence limits associated with these coefficients indicates that they can be regarded as providing very accurate estimates of the internal consistency of the PRMQ in the general adult population. The magnitude of the coefficients demonstrates that the reliability of the PRMQ is very acceptable for use in group studies or work with individuals.

The conversion tables (Tables 2–4) can be used to convert raw scores to *T* scores and obtain confidence limits. As noted, conversion of raw scores on the scales to a common metric facilitates comparison of an individual's scores. A practical example of the use of the tables is presented in the next section. Although we see the conversion tables primarily as an aid to interpretation of scores in work with individuals (i.e., in general medical or mental health services), they could also be usefully employed to set inclusion or exclusion criteria for research purposes. Furthermore, as age and gender did not influence PRMQ scores, the summary statistics presented in Table 1 (i.e., means and *SD*s) could also be used as comparison standards for studies of clinical populations in which a control sample is unavailable.

Reliability and abnormality of discrepancies between prospective and retrospective memory

The meaning and use of the conversion tables and tabled data on the reliability and abnormality (i.e., rarity) of discrepancies between the Prospective and Retrospective scales are best illustrated by a worked example. Take the example of a man referred to a memory clinic with a suspected early stage frontal lobe dementia. His raw scores on the Prospective and Retrospective scales were 32 and 24, respectively. Consulting Table 3, his *T* score on the Prospective scale is 26; this score is therefore over 2 *SD*s below the mean of the normative sample. It can also be seen from Table 3 that the 95% confidence limits on his true score are 24 and 37; there is a 95% probability that his true score lies between these limits. Consulting Table 4, his *T* score on the Retrospective scale is 39 (he is just over 1 *SD* below the normative mean), and the 95% limits are 34 and 48.

From Tables 3 and 4 it can also be seen that his estimated *true* scores on the Prospective and Retrospective scales were 30 and 41 respectively, yielding a discrepancy between true scores of 11 points. Referring to Table 5 it can be seen that this discrepancy exceeds the critical value (10) for significance at the 0.05 level (two-tailed). Therefore, the discrepancy between his Prospective and Retrospective scores is taken to reflect a genuine difference in self-rated memory rather than the effects of measurement error; i.e., it is a reliable difference. One-tailed values for assessing the reliability of difference are also presented in Table 5 as it would be legitimate to use these if the researcher or clinician wished to test a directional hypothesis.

To assess the rarity or abnormality of the discrepancy between his self-rated memory on the Prospective and Retrospective scales we use his *T* scores, *not* his estimated true scores. The discrepancy between *T* scores is 13 points. Referring to Table 6 it can be seen that only 4.4% of the population would be expected to show a discrepancy in favour of Retrospective memory as extreme as that obtained.

The distinction between the *reliability* of a difference and the *abnormality* (or rarity) of a difference is an important one in assessment and the two concepts are often confused (Crawford, 1996; Crawford et al., 1998b). As noted, a reliable difference between the Prospective and Retrospective scales indicates that there is a genuine difference in self-rated memory (i.e., the difference does not simply reflect measurement error). However, many individuals may rate their Prospective memory as better than their Retrospective memory and vice-versa. Therefore, a reliable difference need not be unusual or rare (and, in clinical settings, a reliable difference need not necessarily be a cause for concern). As a result, information on the reliability of a difference should be complemented with information on the rarity of the difference.

Although the labour involved in scoring the PRMQ using the tabulated data presented here is relatively modest, we have prepared a simple computer program (for PCs) to automate scoring and analysis of an individual's PRMQ data. Details of the program and information on how to obtain a copy are presented in Appendix 2.

Potential roles for the PRMQ in clinical practice

It needs to be stressed that the PRMQ is a measure of *self-rated* memory failures and therefore cannot be treated as though it were a direct measure of memory performance. However, there is an increasing recognition in neuropsychology that, where possible, we should employ multiple indicators of the constructs we assess (Crawford, Parker, & McKinlay, 1992). Therefore, the use of formal tests of cognitive functioning should be supplemented with self (and/or proxy) reports and by naturalistic observations etc.

The PRMQ can be used to *quantify* an individual's memory complaints and could be used to identify issues to be followed up in a clinical interview. The PRMQ could also be used as a basis for forming clinical hypotheses that can then be evaluated with direct tests of prospective and retrospective memory. Further, the PRMQ can potentially provide convergent evidence of deficits observed on formal memory tests. Where there is not convergence between formal testing and the PRMQ, this in itself is liable to be useful information. For example, unequivocal evidence of memory deficits on formal testing, coupled with PRMQ scores in the average or above average range, indicates a lack of insight that may be more serious in its consequences than the deficits themselves. The PRMQ is best suited to work in general medical settings and memory clinics. The self-rating version of the PRMQ clearly should not be used in patients with significant comprehension or reading deficits (although, in the latter case, ratings could still be obtained with assistance).

Conclusions and future research

The present study indicates that the PRMQ has a tripartite latent structure; i.e., in addition to measuring general self-rated memory it also measures specific components corresponding to prospective and retrospective memory. This structure appears to be very "clean" in that the fit of the tripartite model was good and the specification of other potential factors did not account for additional covariance among items. Furthermore, the reliabilities of the Total and Prospective and retrospective scales were high. The PRMQ has the potential advantage over other self-report scales in that it balances prospective and retrospective items, and measures these constructs systematically over a range of contexts. The provision of norms means that self-rated memory can be readily quantified (as can the degree of confidence attached to the ratings). The utility of the scale is increased by the provision of methods that allow evaluation of discrepancies between the prospective and retrospective ratings.

The current norms are based on *self*-report of prospective and retrospective memory. It would be very worthwhile to establish similar norms for *proxy* ratings (i.e., ratings by spouses or others). In addition, although the limited existing data on the relationships between the PRMQ and formal memory tests are encouraging, this issue should be further explored and also extended to include the study of variables that may moderate these relationships. Finally, it would also be important to examine the latent structure of the PRMQ in other cultures (the PRMQ has been translated into a number of European languages) and in clinical populations in order to determine the extent to which the scale is factorially invariant.

Manuscript received 21 May 2001
Manuscript accepted 20 December 2001

REFERENCES

Bentler, P.M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P.M., & Mooijart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. *Psychological Bulletin*, *106*, 315–317.

Brandimonte, M., Einstein, G.O., & McDaniel, M.A. (Eds.). (1996). *Prospective memory: Theory and applications*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Broadbent, D.E., Cooper, P.F., Fitzgerald, P., & Parkes, K.R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, *21*, 1–16.

Byrne, B.M. (1994). *Structural equation modeling with EQS and EQS/Windows*. Thousand Oaks, CA: Sage.

Craik, F.I.M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and performance* (pp. 409–422). Amsterdam: Elsevier.

Crawford, J.R. (1996). Assessment. In J.G. Beaumont, P.M. Kenealy & M.J. Rogers (Eds.), *The Blackwell*

dictionary of neuropsychology, (pp. 108–116). London: Blackwell.
Crawford, J.R., Howell, D.C., & Garthwaite, P.H. (1998a). Payne and Jones revisited: Estimating the abnormality of test score differences using a modified paired samples *t*-test. *Journal of Clinical and Experimental Neuropsychology*, *20*, 898–905.
Crawford, J.R., Parker, D.M., & McKinlay, W.W. (Eds.). (1992). *A handbook of neuropsychological assessment*. Hove, UK: Lawrence Erlbaum Associates Ltd.
Crawford, J.R., Venneri, A., & O'Carroll, R.E. (1998b). Neuropsychological assessment of the elderly. In A.S. Bellack & M. Hersen (Eds.), *Comprehensive clinical psychology, Vol. 7: Clinical geropsychology* (pp. 133–169). Oxford: Pergamon.
Dobbs, A.R., & Rule, B.G. (1987). Prospective memory and self reports of memory abilities in older adults. *Canadian Journal of Psychology*, *41*, 209–222.
Ellis, J., & Kvavilashvili, L. (2000). Prospective memory in 2000: Past present and future directions. *Applied Cognitive Psychology*, *14*, S1–S10.
Glutting, J.J., Mcdermott, P.A., & Stanley, J.C. (1987). Resolving differences among methods of establishing confidence limits for test scores. *Educational and Psychological Measurement*, *1987*, 607–614.
Hu, L.T., & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453.
Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariances structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.
Mantyla, T. (2001). *Prospective memory complaint and impairment in middle-aged adults*. Manuscript submitted for publication.
Maylor, E.A. (1996). Does prospective memory decline with age? In M. Brandimonte, G.O. Einstein, & M.A. McDaniel (Eds.), *Prospective memory: Theory and applications* (pp. 173–197). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
Maylor, E.A., Darby, R.J., Logie, R.H., Della Sala, S., & Smith, G. (in press). Prospective memory across the lifespan. In N. Ohta & P. Graf (Eds.), *Lifespan memory development*. Cambridge, MT: MIT Press.
Maylor, E.A., Smith, G., Della Sala, S., & Logie, R.H. (2002). *Prospective and retrospective memory in normal aging and dementia: An experimental study*. Manuscript submitted for publication.
Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd Edn.). New York: McGraw-Hill.
Rabbitt, P., Maylor, E., McInnes, L., Bent, N., & Moore, B. (1995). What goods can self-assessment questionnaires deliver for cognitive gerontology? *Applied Cognitive Psychology*, *9*, S127–S152.
Silverstein, A.B. (1989). Confidence intervals for test scores and significance tests for test score differences: A comparison of methods. *Journal of Clinical Psychology*, *45*, 828–832.
Smith, G.V., Della Sala, S., Logie, R.H., & Maylor, E.A.M. (2000). Prospective and retrospective memory in normal ageing and dementia: A questionnaire study. *Memory*, *8*, 311–321.
Stanley, J.C. (1971). Reliability. In R.L. Thorndike (Ed.), *Educational measurement* (2nd Edn., pp. 356–442). Washington DC: American Council on Education.
Steiger, J.H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser. *Structural Equation Modeling*, *7*, 149–162.
Sunderland, A., Harris, J.E., & Baddeley, A.D. (1984). Assessing everyday memory after severe head injury. In J.E. Harris & P.E. Morris (Eds.), *Everyday memory, actions and absent mindedness*. London: Academic Press.
Sunderland, A., Harris, J.E., & Baddeley, A.D. (1988). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behavior*, *22*, 341–357.
Zacks, R.T., Hasher, L., & Li, K.Z.H. (2000). Human memory. In F.I.M. Craik & T.A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 293–357). Mahwah, NJ: Erlbaum.

APPENDIX 1

PRMQ items and their categorisations

Item No.	*Item*	*Prospective vs Retrospective*	*Short- vs Long-term*	*Self-cued vs Envir. cued*
1	Do you decide to do something in a few minutes' time and then forget to do it?	Prospective	Short-term	Self-cued
2	Do you fail to recognise a place you have visited before?	Retrospective	Long-term	Envir. cued
3	Do you fail to do something you were supposed to do a few minutes later even though it's there in front of you, like take a pill or turn off the kettle?	Prospective	Short-term	Envir. cued
4	Do you forget something that you were told a few minutes before?	Retrospective	Short-term	Self-cued
5	Do you forget appointments if you are not prompted by someone else or by a reminder such as a calendar or diary?	Prospective	Long-term	Self-cued
6	Do you fail to recognise a character in a radio or television show from scene to scene?	Retrospective	Short-term	Envir. cued
7	Do you forget to buy something you planned to buy, like a birthday card, even when you see the shop?	Prospective	Long-term	Envir. cued
8	Do you fail to recall things that have happened to you in the last few days?	Retrospective	Long-term	Self-cued
9	Do you repeat the same story to the same person on different occasions?	Retrospective	Long-term	Envir. cued
10	Do you intend to take something with you, before leaving a room or going out, but minutes later leave it behind, even though it's there in front of you?	Prospective	Short-term	Envir. cued
11	Do you mislay something that you have just put down, like a magazine or glasses?	Retrospective	Short-term	Self-cued
12	Do you fail to mention or give something to a visitor that you were asked to pass on?	Prospective	Long-term	Envir. cued
13	Do you look at something without realising you have seen it moments before?	Retrospective	Short-term	Envir. cued
14	If you tried to contact a friend or relative who was out, would you forget to try again later?	Prospective	Long-term	Self-cued
15	Do you forget what you watched on television the previous day?	Retrospective	Long-term	Self-cued
16	Do you forget to tell someone something you had meant to mention a few minutes ago?	Prospective	Short-term	Self-cued

APPENDIX 2

Computer program for scoring the PRMQ

The computer program takes an individual's raw scores on the Prospective and Retrospective scales as input and provides T scores, estimated true scores, and 95% confidence limits for all three scales. It also tests whether the discrepancy between the Prospective and Retrospective scales is reliably different, and records the estimated percentage of the population that will equal or exceed the observed discrepancy. In addition to being a quick and convenient means of scoring and analysing the PRMQ, the program will also help to avoid clerical errors. The program can be downloaded from the following web site address: www.psyc.abdn.ac.uk/homedir/jcrawford/prmq.htm