Test–Retest Reliability and Random Measurement Error of the Multifactorial Memory Questionnaire in Older Adults With Subjective Memory Complaints

Hui-Ling Yang, PhD[1] [iD]**, Kuei-Ru Chou, PhD, RN[2,3,4],**
Shih-Chieh Lee, PhD[5], Po-Han Lin, PhD[6],
and Hsin-Yu Chiang, PhD[1]

Abstract

To examine the psychometric properties of the Multifactorial Memory Questionnaire (MMQ) in older adults with subjective memory complaints. The three MMQ subscale (Satisfaction, Ability, and Strategy) was administered twice, with a 3-month interval. The test-retest reliability was examined using intraclass correlation coefficients (ICCs). The random measurement error was examined by calculating the standard error of measurement (SEM) and minimal detectable change (MDC_{95}). The test-retest reliabilities of the three MMQ subscales were generally acceptable. The SEM of the three MMQ subscales was higher than the acceptable criterion of 10%. Despite the influence of random measurement error, the change scores of the three MMQ subscales may represent true changes if they are larger than the MDC_{95} of 13.2 (Satisfaction), 18.4 (Ability), and 16.9 (Strategy). The MMQ appears to be a reliable measure for use in research settings, but may not yet be suitable for clinical use.

Introduction

Subjective memory complaints (SMC) reflect an individual's dissatisfaction with his or her memory function, which is common in community-dwelling older adults (Matthews et al., 2020; Mitchell et al., 2014). SMC appears to be associated with poor outcomes, including reduced social participation, increased emotional symptoms (e.g., depression and anxiety), and low quality of life (Mogle et al., 2017; Montejo et al., 2011). Moreover, SMC reportedly is a risk factor for further functional decline, including cognitive decline and increased experienced difficulties in performing activities of daily living (ADLs; Cordier et al., 2019). Subjective memory assessment can help clinicians detect early stages of cognitive decline in older adults with SMC. Interpreting change is a requisite component of clinical decision-making for clinicians. Moreover, to interpret the measurement results, clinicians have to decide whether the change scores in the subjective memory measures are true or beyond measurement error (Seamon et al., 2022). Therefore, a reliable and clinically applicable SMC assessment tool that allows the assessor to effectively distinguish whether the subject has SMC is essential.

The Multifactorial Memory Questionnaire (MMQ; Troyer & Rich, 2002) is among the most commonly used measures for assessing self-perceived memory function in older adults (Troyer et al., 2019). It appears

[1]Fu Jen Catholic University, New Taipei City
[2]Taipei Medical University
[3]Taipei Medical University-Shuang Ho Hospital
[4]Taipei Medical University Hospital
[5]National Taiwan University, Taipei
[6]National Taipei University of Education

Corresponding Author:
Hui-Ling Yang, Department of Nursing, College of Medicine, Fu Jen Catholic University, No. 510, Zhongzheng Road, Xinzhuang District, New Taipei City, 242.
Email: 151276@mail.fju.edu.tw

to be a promising measure for two reasons. First, the MMQ includes an independent dimension (i.e., the satisfaction subscale) to assess examinees' subjective feelings about memory (Troyer & Rich, 2002). SMC is linked to subjective feelings about memory (Hanninen et al., 1994; Lee, 2016); specifically, SMC may have a direct negative impact on emotional symptoms in older adults (Kawagoe et al., 2019; Schweizer et al., 2018). Thus, the MMQ provides more information for assessing SMC in addition to the "ability" (e.g., the frequency of making memory mistakes) and/or "strategy" (e.g., frequency of using memory strategies and aids) dimensions included in most measures. Second, in healthy older adults, the MMQ has shown preliminary good psychometric properties, including test-retest reliability (Simon et al., 2016; Troyer & Rich, 2002), convergent validity (Troyer & Rich, 2002; van der Werf & Vos, 2011), and random measurement error (Troyer et al., 2019). Accordingly, the MMQ has great potential to be used as a measure for assessing older adults with SMC.

Reliability refers to the stability of the scores provided by a measure. It is considered a basic psychometric property of all measures; thus, reliability validation is a pre-requisite of all psychometric validations. The reliability validation generally includes test-retest reliability and the amount of random measurement error. Test-retest reliability refers to the degree of consistency of observed scores in repeated assessments (Schambra et al., 2015). Random measurement error is the variation of observed scores caused by random factors (Mellenbergh, 2019). Because these two indices provide different information about MMQ score reliability, reliability validation should ideally include both test-retest reliability and random measurement error simultaneously.

However, the MMQ has not been validated in older adults with SMC. Particularly, because the psychometric properties are sample dependent (Hobart & Cano, 2009), the previous psychometric evidence on healthy older adults (Troyer & Rich, 2002) cannot be directly inferred to the SMC population. This lack of psychometric evidence may interfere with the MMQ score interpretations in older adults with SMC. Furthermore, from a previous review article on the psychometric properties of MMQ (Troyer et al., 2019), the minimal detectable change (MDC_{95}) of the MMQ has not been examined, which may thus limit the clinical applicability of this tool (Seamon et al., 2022).

Accordingly, validation of the psychometric properties of the MMQ is essential to determine the utility of the MMQ in the SMC population. The present study aimed to investigate additional psychometric properties of the MMQ in older adults with SMC: test-retest reliability, random measurement error, heteroscedasticity, and systematic bias.

Methods

This study used a qualitative psychometric, repeated-assessments design (with a 3-month interval) and adhered to the STROBE guideline.

Study Design and Participants

The data were extracted from a research project that investigated the effect of cognitive function training in older adults with SMC; some data have been published in a previous study (Yang et al., 2019). We recruited older adults (age ≥65 years) by convenience sampling mainly from senior centers and homes. The inclusion criteria were as follows: (1) SMC, reports of at least two or more memory failures (e.g., forgetting an appointment or someone's name; Silva et al., 2014); (2) Montreal Cognitive Assessment (MoCA) score >17 points; (3) ability to perform ADL independently (assessed using the Barthel Index and Lawton's instrumental activities of daily living scales; Jekel et al., 2015); and (4) can understand and follow written instructions. The participants were excluded if they had major neurocognitive disorders (e.g., dementia) or a rapid decline in cognitive or physical function. The study was approved by the Joint Institutional Review Board. Informed consent was obtained from the participants (approval no.:201301045).

Procedure

The MMQ was tested in three senior centers and homes throughout Taiwan from 2015 to 2017. Specifically, MMQ data were extracted at two-time points during the follow-up period of the previous study (i.e., 3- and 6-month follow-up; Yang et al., 2019). In addition, the data on individual demographics, ADL (i.e., Barthel Index and IADL scales), MoCA, and MMQ (baseline data) scores were extracted. To avoid the interference of external factors on the participants' judgments, assessments were performed in a quiet room under the same conditions and environment throughout the study. The MMQ was filled in by the participants or with the help of a research assistant. All other information was provided by the participants.

Measures

The MMQ (Troyer & Rich, 2002) consists of 57 items that measure the individuals' self-perceived problems of memory function over the past 2 weeks using a 5-point Likert-type scale (0–4). These items are categorized into three subscales (i.e., Satisfaction, Ability, and Strategy). The Satisfaction subscale (formerly called *contentment*) contains 18 items assessing the satisfaction and feelings related to one's memory (scores range from 0 to 72). The Ability subscale contains 20 items assessing the self-perceptions of everyday memory ability (scores range from 0 to 80). The Strategy subscale contains 19 items assessing the frequency of using memory strategies and aids in everyday life (scores range from 0 to 76). Higher scores indicate greater self-reported memory function for the three MMQ subscales. The MMQ takes approximately 10 to 15 minutes to complete (Troyer & Rich, 2018).

The MoCA is a screening tool for cognitive impairment that assesses multiple cognitive domains (e.g., visuospatial/executive function and memory). The scores range from 0 to 30, with higher scores indicating greater cognitive function. A score of ≤17 indicates cognitive impairment (Tsai et al., 2016). The MoCA takes approximately 10 minutes to complete (Nasreddine et al., 2005).

ADL assessment is performed using the Barthel Index and Lawton's IADL (Bouwstra et al., 2019; Lawton & Brody, 1969). The Barthel Index assessment includes 10 items assessing basic self-maintenance skills, such as bathing, dressing, or eating. The scores range from 0 to 100, with higher scores indicating greater independence. Lawton's IADL assessment consists of eight items for assessing more complex activities, such as using public transportation, managing finances, or shopping. The scores range from 0 to 8, with higher scores indicating greater independence.

Data Analysis

Test-Retest Reliability. We calculated the intraclass correlation coefficients (ICCs) based on a two-way random model with an absolute agreement (2,1; Qin et al., 2019; Weir, 2005). An ICCs value of ≥.90 was considered to indicate excellent reliability; .75 to .90, good reliability; .50 to.75, moderate reliability; and <.50, poor reliability (Koo & Li, 2016).

Random Measurement Error. We examined the MMQ's random measurement error by standard error of measurement (SEM). Moreover, to improve the subscale scores' interpretability given the influence of random measurement error, the MDC_{95} was also calculated.

The SEM is a measure of the degree to which observed scores are spread around a true score, which can be calculated based on the ICCs value. Because the SEM is dependent of the unit of measurement (Hars et al., 2013), it was expressed as a percentage of the mean (i.e., SEM%) to produce a unitless indicator and allow for comparison across different measures (Hars et al., 2013). The SEM and SEM% formulas are as follows:

$$SEM = SD_{Baseline} \times \sqrt{(1-r)}$$
$$SEM\% = (SEM/mean) \times 100$$

In these formulas, $SD_{Baseline}$ is the standard deviation (SD) of the baseline scores and r is the test-retest reliability coefficient estimated from ICCs. Moreover, the *mean* represents the average of all test and retest scores. A SEM% <10% was considered acceptable (Braun et al., 2019).

The MDC_{95} is a threshold for determining whether the change score between consecutive assessments exceeds a true change with a 95% confidence interval. The MDC_{95} values were calculated based on the SEM value using the following formula (Haley & Fragala-Pinkham, 2006):

$$MDC_{95} = SEM \times 1.96 \times \sqrt{2}$$

In this formula, the value of 1.96 was the z-score corresponding to the 95% confidence interval. The square root of 2 was used to adjust for sampling from repeated assessments. A smaller MDC_{95} indicates greater reliability (Haley & Fragala-Pinkham, 2006).

Heteroscedasticity. Heteroscedasticity was examined by Bland–Altman plot and Pearson's correlation coefficient (r) between the means and absolute differences in *consecutive* assessments.

The Bland–Altman plot was used to reveal whether the discrepancies between the two assessments were distributed randomly and equally across individuals with diverse levels of self-perceived memory function (i.e., heteroscedasticity). This plot is constructed by plotting the means of *consecutive* assessments (X axis, i.e., the individuals' levels of self-perceived memory function) against the differences (Y axis, i.e., the consistency of two assessments) among *consecutive* assessments (Bland & Altman, 1986; Doğan, 2018). The limit of agreement (LOA; the range with approximately 95% of the data) is presented in the Bland–Altman plot to identify possible outliers in the data (Bland & Altman, 1986; Doğan, 2018). Furthermore, we used Pearson's correlation coefficient (r) to examine whether heteroscedasticity existed between the means and absolute differences of *consecutive* assessments. An absolute value of $r > .30$ was considered heteroscedastic (Atkinson & Nevill, 1998).

Systematic Bias

Systematic bias was examined by paired *t*-test and effect size. The differences between the groups were considered statistical significance if the *p* values of .05 or lower, indicating systematically biased. The effect size was calculated by dividing the difference between *the two assessment means* by the pooled *SD* (i.e., root mean square of the two *SDs*). The effect size *d* was interpreted with $0.20 > d \leq 0.49$ being defined as small, $0.50 \leq d \leq 0.79$ being defined as moderate, and $d \geq 0.80$ being defined as large (Nascimento Barros et al., 2018).

Data were analyzed using the IBM Statistical Package for Social Science (SPSS) version 22.0 software (SPSS Inc., Chicago, IL, USA).

Results

Altogether, 101 participants with mean age close to 80 years were recruited and completed all study

assessments. More than two-thirds of the participants were female. The participants' characteristics are listed in Table 1.

Table 2 shows the test-retest reliability and random measurement error of the three MMQ subscales. The ICCs of the MMQ Satisfaction, Ability, and Strategy subscales were 0.72, 0.74, and 0.77, respectively. The SEM% were 11.7%, 11.2%, and 28.2%, respectively, which were higher than the acceptable criterion of 10%. The MDC_{95} of the three MMQ subscales was 13.2 to 18.4.

Table 1. Demographic Characteristics of the Participants ($n = 101$).

Variable	Value
Age (years), mean $\pm$ SD	78.8 $\pm$ 7.9
Sex, n (%)	
Male	23 (22.8)
Female	78 (77.2)
Education (years), mean $\pm$ SD	10.8 $\pm$ 4.8
BADL, mean $\pm$ SD	98.0 $\pm$ 9.7
IADL, mean $\pm$ SD	7.7 $\pm$ 1.1
MoCA, mean $\pm$ SD	22.5 $\pm$ 5.3
MMQ satisfaction, mean $\pm$ SD	41.3 $\pm$ 8.6
MMQ ability, mean $\pm$ SD	41.3 $\pm$ 8.6
MMQ strategy, mean $\pm$ SD	22.3 $\pm$ 12.1

Note. SD = standard deviation; BADL = basic activities of daily living; IADL = instrumental activities of daily living; MoCA = montreal cognitive assessment.

The Bland–Altman plots (Figure 1) showed that discrepancy distribution between the two assessments of the three subscales were generally random. Moreover, the amounts of discrepancy were equal across individuals with diverse levels of self-perceived memory function. The upper and lower LOAs were −15.7 and 22.0 points, respectively, for the MMQ Satisfaction subscale, −24.2 and 24.7 points for the MMQ Ability subscale, and −21.4 and 21.3 points for the MMQ Strategy subscale.

Pearson's *r* values for the associations between the means and absolute differences for the MMQ Satisfaction, Ability, and Strategy subscales were −0.08, −0.16, and 0.16 respectively.

Regarding the systematic bias (Table 2), scores of two MMQ subscales were not significantly different across the two assessments (ES = 0.01 and −0.02, *p*s = .855 and .949 for the Ability and Strategy subscales, respectively), but that of the Satisfaction subscale was (ES = 0.34, *p* = .001).

Discussion

Test-Retest Reliability

We found that the ICCs values of the three MMQ subscales ranged from .72 to .77, indicating acceptable test-retest reliability of MMC in older adults with SMC for group comparisons. These findings are similar to those

Table 2. Test–Retest Reliability and Random Measurement Error of the MMQ (Raw Scores; $n = 101$).

MMQ subscale[a]	First ($M \pm SD$)	Second ($M \pm SD$)	Mean difference	Paired *t* test (*p*-value)	Effect size	ICC [95% CI]	SEM	SEM%	MDC_{95}
Satisfaction	40.6 $\pm$ 9.0	43.8 $\pm$ 11.3	3.1	3.28 (.001)	0.34	0.72	4.8	11.7%	13.2
Ability	59.1 $\pm$ 13.0	59.3 $\pm$ 14.6	0.2	0.18 (.855)	0.01	0.74	6.6	11.2%	18.4
Strategy	21.6 $\pm$ 12.7	21.5 $\pm$ 12.5	−0.1	−0.64 (.949)	−0.02	0.77	6.1	28.2%	16.9

Note. MMQ = the Multifactorial Memory Questionnaire; SD = standard deviation; CI = confidence interval; ICC = intraclass correlation coefficient; SEM = standard error of measurement; MDC_{95} = minimal detectable change.
[a]Total score range of the MMQ Satisfaction subscale, 0 to 72 points; Total score range of the MMQ Ability subscale, 0 to 80 points; Total score range of the MMQ Strategy subscale, 0 to 76 points.
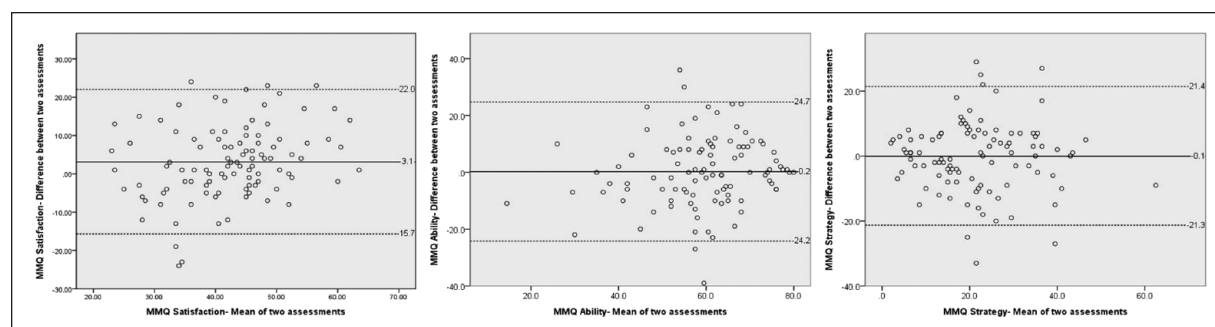

Figure 1. Bland–Altman plots showing the means (*X* axis) of the *consecutive* assessments against the differences (*Y* axis) between *consecutive* assessments of 3 MMQ subscales.
Note. The solid central horizontal line shows the mean differences in consecutive assessments. The two dashed horizontal lines represent 95% limits of agreement (mean difference $\pm$ 1.96 $\times$ $SD_{difference}$).

of a previous study (ICCs = .74–.80 for healthy older adults with a 3-month interval; Simon et al., 2016), suggesting that the scores seem reliable in repeated assessments. Given that an ICCs of .70 is considered acceptable for group-level comparisons, the MMQ has sufficient reliability for use in research settings (e.g., comparisons targeting the differences in mean scores between the intervention and control groups). However, because the MMQ ICCs were < .90 (minimum requirement for individual comparison), the MMQ may not be sufficiently reliable for clinical settings (e.g., evaluating individual differences in older adults; Braun et al., 2019; Hafner et al., 2016). Therefore, MMQ users should consider the purpose of use (group-level or individual-level) and interpret the scores appropriately.

Random Measurement Error

The SEM% of two MMQ subscales was slightly higher than the acceptance criterion of 10% (SEM% of the Satisfaction and Ability subscales were 11.7% and 11.2%, respectively), but that of the Strategy subscale (28.2%) was much higher. These findings indicate that the random measurement errors of the MMQ may not be satisfactory, which may partly explain the modest ICCs mentioned earlier (i.e., scores had been affected by random measurement errors, resulting in lowered test-retest reliability). Among the three subscales, the SEM% of the Strategy subscale was higher than those of the other two subscales. These findings are similar to those in previous studies on healthy older adults (SEM% of the Satisfaction, Ability, and Strategy subscales were 12.6%, 8.2%, and 20.9%, respectively; Simon et al., 2016; Troyer et al., 2019), suggesting that the larger SEM% of the Strategy subscale may not be unique to older adults with SMC. One possible explanation for this may be that the Strategy subscale's mean score was lower (pre-test: 21.6) than the other two scales' scores (pre-test: 40.6 and 59.1 for the Satisfaction and Ability subscales, respectively). Given that a smaller mean score is very likely to yield a larger SEM%, the current SEM% of the Strategy subscale may not be overestimated. To resolve this problem, future studies can include individuals using more memory strategies or consider revising the Strategy subscale (e.g., adding some easier items).

We found that the MDC_{95} values of the three MMQ subscales were 13.2 (Satisfaction), 18.4 (Ability), and 16.9 (Strategy), indicating the lowest score for true change in each domain. Such values are useful for both clinicians and researchers to interpret individuals' changes in self-reported memory functions. For clinicians, the MDC_{95} can be used to determine whether the individual changes are beyond random measurement errors (Atkinson & Nevill, 1998; Mellenbergh, 2019). For instance, a Satisfaction subscale score change exceeding 14 points between two assessments can be regarded as a true change. For researchers, the percentage of individuals whose change scores exceed the

MDC_{95} can be calculated as an index of the effectiveness of the intervention (i.e., the higher the percentage, the larger the effects). Accordingly, the MDC_{95} can be helpful in interpreting the MMQ's change scores, given the influences of random measurement error on scoring in both clinical and research settings.

Heteroscedasticity

Regarding heteroscedasticity, the Bland–Altman plot showed that the distributions of the discrepancies between two assessments of the three subscales were generally random. Moreover, the Pearson's *r* values between the absolute differences and means of the three MMQ subscales were < .30 (−.08, −.16, and .16 for the Satisfaction, Ability, and Strategy subscales, respectively). These results suggest that the amount of random measurement error was consistent for an individual with different levels of self-reported memory functions; therefore, heteroscedasticity was not present. Accordingly, the fixed, aforementioned MDC_{95} values can be used for all individuals to determine whether a difference between *consecutive* assessments reflects a true change beyond random measurement error, regardless of the individuals' MMQ scores.

Systematic Bias

Our results showed no statistically significant differences among the test–retest assessments in either the MMQ Ability (*p* = .855, ES = 0.01) or Strategy (*p* = .949, ES = −0.02) subscales. These findings indicate that the individuals' scores did not systematically change across a 3-month interval, suggesting that these two subscales have no systematic bias. Contrarily, for the MMQ Satisfaction subscale, a significant difference (paired *t*-test, *p* = .001, ES = 0.34) was found, implying that the score systematically improved. However, given that the amount of the difference in the Satisfaction subscale was relatively small (3.1 points, approximately 4.3% of the total score), systematic bias may not be a major concern for clinicians who repeatedly administer the MMQ. Accordingly, the MMQ scores appear to be unbiased.

Study Limitations

This study has two limitations. First, convenience sampling was used. Second, the interval between repeated assessments was long (i.e., 3 months). These two issues may limit the findings' generalizability. Future studies are needed to confirm the MMQ's validity in older adults with SMC.

Conclusions and Implications

The MMQ appears to be a reliable measure with acceptable test-retest reliability for use in research settings.

However, the random measurement errors of all the MMQ scales were higher than the acceptable criterion. Thus, the MMQ may be more appropriate for research purposes than for clinical purposes.

Clinical Implication and Future Studies

Despite some uncertainty about the psychometric properties of the MMQ subscales, we believe that the results of this study are reliable enough to recommend that researchers apply it to memory training in older adults to assess self-reported memory complaints during the interventions. The multidimensional assessment features of the MMQ can be used to quantify the experience of older adults with SMC in terms of strategies, ability, and satisfaction. Theoretically, MMQ could provide insights into participants' perceptions of the progression of self-memory changes, which could be a starting point for shared decision-making and goal-setting in memory-related interventions.

To some extent, the assessment of SMC can be used as a predictor of cognitive function or ADL degradation and can provide a clinically valuable reference range. Therefore, the use of MMQ can help clinical professionals to identify barriers that may hinder the cognitive decline of older adults, enhance the sensitivity of clinical care, and thereby maintain or improve the cognitive function of older adults, providing the most effective individualized care. Additionally, MMQ was originally developed and validated for use with middle-aged and older adults in the community-dwelling, but may also be useful for other older adults outside the community-dwelling or those with memory/cognitive impairments. Further research could further examine the psychometric properties of MMQ in other cognitive impairment populations is recommended.

ORCID iD

Hui-Ling Yang https://orcid.org/0000-0003-2322-4157

References

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, *26*(4), 217–238. https://doi.org/10.2165/00007256-199826040-00002

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bouwstra, H., Smit, E. B., Wattel, E. M., van der Wouden, J. C., Hertogh, C. M., Terluin, B., & Terwee, C. B. (2019). Measurement properties of the Barthel Index in geriatric rehabilitation. *Journal of the American Medical Directors Association*, *20*(4), 420–425. https://doi.org/10.1016/j.jamda.2018.09.033

Braun, T., Ehrenbrusthoff, K., Bahns, C., Happe, L., & Kopkow, C. (2019). Cross-cultural adaptation, internal consistency, test-retest reliability and feasibility of the German version of the evidence-based practice inventory. *BMC Health Services Research*, *19*(1), 455. https://doi.org/10.1186/s12913-019-4273-0

Cordier, R., Chen, Y. W., Clemson, L., Byles, J., & Mahoney, N. (2019). Subjective memory complaints and difficulty performing activities of daily living among older women in Australia. *Australian Occupational Therapy Journal*, *66*(2), 227–238. https://doi.org/10.1111/1440-1630.12548

Doğan, N. Ö. (2018). Bland-Altman analysis: A paradigm to understand correlation and agreement. *Turkish Journal of Emergency Medicine*, *18*(4), 139–141. https://doi.org/10.1016/j.tjem.2018.09.001

Hafner, B. J., Morgan, S. J., Askew, R. L., & Salem, R. (2016). Psychometric evaluation of self-report outcome measures for prosthetic applications. *Journal of Rehabilitation Research and Development*, *53*(6), 797–812. https://doi.org/10.1682/JRRD.2015.12.0228

Haley, S. M., & Fragala-Pinkham, M. A. (2006). Interpreting change scores of tests and measures used in physical therapy. *Physical Therapy*, *86*(5), 735–743. https://doi.org/10.1093/ptj/86.5.735

Hanninen, T., Reinikainen, K. J., Helkala, E. L., Koivisto, K., Mykkanen, L., Laakso, M., Pyorala, K., & Riekkinen, P. J. (1994). Subjective memory complaints and personality traits in normal elderly subjects. *Journal of the American Geriatrics Society*, *42*(1), 1–4. https://doi.org/10.1111/j.1532-5415.1994.tb06064.x

Hars, M., Herrmann, F. R., & Trombetti, A. (2013). Reliability and minimal detectable change of gait variables in community-dwelling and hospitalized older fallers. *Gait & Posture*, *38*(4), 1010–1014. https://doi.org/10.1016/j.gaitpost.2013.05.015

Hobart, J., & Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: The role of new psychometric methods. *Health Technology Assessment*, *13*(12), 1–177. https://doi.org/10.3310/hta13120

Jekel, K., Damian, M., Wattmo, C., Hausner, L., Bullock, R., Connelly, P. J., Dubois, B., Eriksdotter, M., Ewers, M., Graessel, E., Kramberger, M. G., Law, E., Mecocci, P., Molinuevo, J. L., Nygård, L., Olde-Rikkert, M. G., Orgogozo, J.-M., Pasquier, F., Peres, K., Salmon, E., Sikkes, S. A., Sobow, T., Spiegel, R., Tsolaki, M., Winblad, B., & Frölich, L. (2015). Mild cognitive impairment and

deficits in instrumental activities of daily living: A systematic review. *Alzheimer's Research & Therapy*, *7*(1), 17. https://doi.org/10.1186/s13195-015-0099-0

Kawagoe, T., Onoda, K., & Yamaguchi, S. (2019). Subjective memory complaints are associated with altered resting-state functional connectivity but not structural atrophy. *NeuroImage: Clinical*, *21*, 101675. https://doi.org/10.1016/j.nicl.2019.101675

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *The Gerontologist*, *9*(3), 179–186.

Lee, P. L. (2016). A joyful heart is good medicine: Positive affect predicts memory complaints. *The American Journal of Geriatric Psychiatry*, *24*(8), 662–670. https://doi.org/10.1016/j.jagp.2016.04.001

Matthews, M. L., Pike, K. E., Wright, B. J., & Kinsella, G. J. (2020). Predictors of long-term improvement in subjective everyday memory following a memory group program for older adults. *Journal of Aging and Health*, *32*(3–4), 216–226. https://doi.org/10.1177/0898264318814997

Mellenbergh, G. J. (2019). Random measurement error. In *Counteracting methodological errors in behavioral research* (pp. 83–106). Springer. https://doi.org/10.1007/978-3-030-12272-0_7

Mitchell, A. J., Beaumont, H., Ferguson, D., Yadegarfar, M., & Stubbs, B. (2014). Risk of dementia and mild cognitive impairment in older people with subjective memory complaints: Meta-analysis. *Acta Psychiatrica Scandinavica*, *130*(6), 439–451. https://doi.org/10.1111/acps.12336

Mogle, J. A., Hill, N., & McDermott, C. (2017). Subjective memory in a national sample: Predicting psychological well-being. *Gerontology*, *63*(5), 460–468. https://doi.org/10.1159/000466691

Montejo, P., Montenegro, M., Fernandez, M. A., & Maestu, F. (2011). Subjective memory complaints in the elderly: Prevalence and influence of temporal orientation, depression and quality of life in a population-based study in the city of Madrid. *Aging & Mental Health*, *15*(1), 85–96. https://doi.org/10.1080/13607863.2010.501062

Nascimento Barros, L. A., Ferrari-Piloni, C., Miranda Torres, É., Estrela, C., & Valladares-Neto, J. (2018). Effect size: A statistical basis for clinical practice. *Revista Odonto Ciencia*, *33*(1), 84–90. https://doi.org/10.15448/1980-6523.2018.1.29437

Nasreddine, Z. S., Phillips Na Fau, Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Qin, S., Nelson, L., McLeod, L., Eremenco, S., & Coons, S. J. (2019). Assessing test–retest reliability of patient-reported outcome measures using intraclass correlation coefficients: Recommendations for selecting and documenting the analytical formula. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *28*(4), 1029–1033. https://doi.org/10.1007/s11136-018-2076-0

Schambra, H. M., Ogden, R. T., Martínez-Hernández, I. E., Lin, X., Chang, Y. B., Rahman, A., Edwards, D. J., & Krakauer, J. W. (2015). The reliability of repeated TMS measures in older adults and in patients with subacute and chronic stroke. *Frontiers in Cellular Neuroscience*, *9*, 335. https://doi.org/10.3389/fncel.2015.00335

Schweizer, S., Kievit, R. A., Emery, T., Cam, C. A. N., & Henson, R. N. (2018). Symptoms of depression in a large healthy population cohort are related to subjective memory complaints and memory performance in negative contexts. *Psychological Medicine*, *48*(1), 104–114. https://doi.org/10.1017/S0033291717001519

Seamon, B. A., Kautz, S. A., Bowden, M. G., & Velozo, C. A. (2022). Revisiting the concept of minimal detectable change for patient-reported outcome measures. *Physical Therapy*, *102*(8), pzac068. https://doi.org/10.1093/ptj/pzac068

Silva, D., Guerreiro, M., Faria, C., Maroco, J., Schmand, B. A., & Mendonça, A. d. (2014). Significance of subjective memory complaints in the clinical setting. *Journal of Geriatric Psychiatry and Neurology*, *27*(4), 259–265. https://journals.sagepub.com/doi/pdf/10.1177/0891988714532018

Simon, S. S., Avila, R. T., Vieira, G., & Bottino, C. M. C. (2016). Metamemory and aging: Psychometric properties of the Brazilian version of the Multifactorial Memory Questionnaire for elderly. *Dementia & Neuropsychologia*, *10*(2), 113–126. https://doi.org/10.1590/S1980-5764-2016DN1002007

Troyer, A. K., & Rich, J. B. (2002). Psychometric properties of a new metamemory questionnaire for older adults. *Journals of Gerontology Series B: Psychological Sciences*, *57*(1), 19–27. https://doi.org/10.1093/geronb/57.1.P19.

Troyer, A. K., & Rich, J. B. (2018). *Multifactorial Memory Questionnaire professional manual*. Baycrest Centre for Geriatric Care. www.baycrest.org/mmq

Troyer, A. K., Leach, L., Vandermorris, S., & Rich, J. B. (2019). The measurement of participant-reported memory across diverse populations and settings: A systematic review and meta-analysis of the Multifactorial Memory Questionnaire. *Memory*, *27*(7), 931–942. https://doi.org/10.1080/09658211.2019.1608255

Tsai, J. C., Chen, C. W., Chu, H., Yang, H. L., Chung, M. H., Liao, Y. M., & Chou, K. R. (2016). Comparing the sensitivity, specificity, and predictive values of the Montreal cognitive assessment and mini-mental state examination when screening people for mild cognitive impairment and dementia in Chinese population. *Archives of Psychiatric Nursing*, *30*(4), 486–491. https://doi.org/10.1016/j.apnu.2016.01.015

van der Werf, S. P., & Vos, S. H. (2011). Memory worries and self-reported daily forgetfulness: A psychometric evaluation of the Dutch translation of the Multifactorial Memory Questionnaire. *Clinical Neuropsycholgist*, *25*(2), 244–268. https://doi.org/10.1080/13854046.2010.543290

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength & Conditioning Research*, *19*(1), 231–240. https://doi.org/10.1519/15184.1

Yang, H. L., Chu, H., Kao, C. C., Chiu, H. L., Tseng, I. J., Tseng, P., & Chou, K. R. (2019). Development and effectiveness of virtual interactive working memory training for older people with mild cognitive impairment: A single-blind randomised controlled trial. *Age and Ageing*, *48*(4), 519–525. https://doi.org/10.1093/ageing/afz029