*The Spanish Journal of Psychology* (2026), **29**, e5, 1–7
doi:10.1017/SJP.2026.10021

**Article**

# Convergent and Discriminant Validity Evidence of the Methodological Quality Scale for Studies Based on Observational Methodology (MQSOM)

Daniel López-Arenas[1], Susana Sanduvete-Chaves[1], José Mena-Raposo[1] and Salvador Chacón-Moscoso[1,2]

[1]Universidad de Sevilla, Spain and [2]Universidad Autónoma de Chile, Chile

**Abstract**

The use of observational methodology has become increasingly more common in psychological research, highlighting the need for tools that ensure methodological rigor. This study presents evidence of convergent/discriminant validity for the Methodological Quality Scale for Studies Based on Observational Methodology (MQSOM). A multitrait-multimethod (MTMM) analysis with Spearman's correlations was used to examine the relationship between MQSOM dimensions and those of three instruments: the Methodological Rigor in Mixed Methods (MRMM), the Guidelines for Reporting Evaluations Based on Observational Methodology (GREOM), and the Mixed Methods Appraisal Tool (MMAT). Ninety-six articles were coded using MQSOM and the instruments for comparison. The MQSOM's design converged with the MRMM's mixed-methods design ($\rho = .217$, $p = .034$), GREOM's design ($\rho = .217$, $p = .034$), and MMAT's qualitative (QUAL) component ($\rho = .212$, $p = .038$). The MQSOM's measurement and analysis aligned with MRMM's data analysis ($\rho = .611$, $p < .001$), GREOM's data quality control ($\rho = .423$, $p < .001$) and results ($\rho = .328$, $p = .001$), and MMAT's quantitative (QUANT) ($\rho = .214$, $p = .037$) and mixed-methods ($\rho = .643$, $p < .001$) components. MQSOM's design exhibited discriminant validity from MRMM's data collection ($\rho = .025$, $p = .807$) and data analysis ($\rho = -.051$, $p = .620$), GREOM's data quality control ($\rho = .025$, $p = .812$) and results ($\rho = -.032$, $p = .759$), and MMAT's QUANT component ($\rho = -.035$, $p = .733$). This study reinforces the validity of MQSOM as a useful methodological quality scale.

**Keywords:** convergent-discriminant validity evidence; Methodological Quality Scale for Studies Based on Observational Methodology (MQSOM); mixed methods; multitrait-multimethod

(Received: 09 July 2025; revised: 03 November 2025; accepted: 19 December 2025)

The use of the observational methodology in study designs allows the spontaneous behaviors of humans in natural contexts to be recorded and quantified (Anguera et al., 2020). Such methods are frequently employed in psychological research as they require minimal intervention and are independent from standardized measurement tools, thus enabling the collection of unbiased information (Anguera et al., 2018).

The observational methodology described above differs from that used in the observational studies typically found in health research. These studies are classified by quantitative (QUANT) design type, including cohort, case–control, and cross-sectional studies. In these studies, empirical groups are compared to identify cause–effect relationships when the feasibility of randomization and experimental control is limited (Cochran & Chambers, 1965). In contrast, observational methodology is considered a mixed-methods approach (Anguera et al., 2017), given that it transforms and connects both qualitative (QUAL) and QUANT elements (Cresswell & Plano-Clark, 2017). Specifically, research designs based on observational methodology can be organized in a QUAL–QUANT–QUAL sequence. The initial QUAL phase involves the study design, the choice of the observation instrument, and the systematic recording of QUAL data. The subsequent QUANT phase focuses on transforming the QUAL data into a code matrix and conducting data quality control procedures prior to data analysis. Finally, the last QUAL phase involves interpreting the results within the context of the initial problem and previous literature (Anguera et al., 2020).

Over the past decade, the increase in studies based on observational methodology has prompted numerous researchers to direct their attention to the quality concerns associated with this mixed-methods approach (Chacón-Moscoso et al., 2013, 2019; Portell et al., 2015). This problem could be addressed by ensuring homogeneity and transparency in both studies that rely on observational methodology and in the criteria established for assessing their quality (Chacón-Moscoso et al., 2013).

In order to achieve this, the Methodological Quality Scale for Studies Based on Observational Methodology (MQSOM) was developed (Chacón-Moscoso et al., 2019; Sanduvete-Chaves et al., 2025),

**Corresponding author:** Susana Sanduvete-Chaves; Email: sussancha@us.es

**Cite this article:** López-Arenas, D., Sanduvete-Chaves, S., Mena-Raposo, J., & Chacón-Moscoso, S. (2026). Convergent and Discriminant Validity Evidence of the Methodological Quality Scale for Studies Based on Observational Methodology (MQSOM). *The Spanish Journal of Psychology* **29**, e5, 1–7. https://doi.org/10.1017/SJP.2026.10021

© The Author(s), 2026. Published by Cambridge University Press on behalf of Universidad Complutense de Madrid and Colegio Oficial de la Psicología de Madrid. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use and/or adaptation of the article.

drawing on the Guidelines for Reporting Evaluations Based on Observational Methodology (GREOM) (Portell et al., 2015). The MQSOM is comprised of 11 items that assess a second-order dimension of global methodological quality through two first-order dimensions: the quality of design and the quality of measurement and analysis. It specifies the minimum methodological aspects necessary to improve the quality of studies based on observational methodology. Adequate reliability and validity evidence based on its internal structure was obtained.

The main objective of this study was to obtain empirical evidence of the convergent and discriminant validity of the MQSOM when compared with other instruments that measure methodological quality in mixed-methods studies.

## Method

### *Participants (Units of Analysis)*

The 96 papers examined for the validation of the MQSOM were obtained from among 650 studies found in an exhaustive search completed for a previous investigation (Sanduvete-Chaves et al., 2025). All met the following inclusion criteria: (a) The observational methodology was used, (b) the work was original, (c) the publication presented the usual sections of a research article (introduction, method, results, and discussion), and (d) it was written in English or Spanish. Consequently, this study employed stratified random sampling to guarantee a homogeneous distribution of the units of analysis across the range of MQSOM scores. The required sample size was calculated using G*Power for 90% power and a 5% type I error rate to detect a standardized effect size of 0.4 or larger, resulting in a sample size of 96 papers.[1]

### *Instruments*

The MQSOM (Sanduvete-Chaves et al., 2025) was used for this study. The scale has adequate psychometric properties (root-mean-square error of approximation [RMSEA] = .000; non-normed fit index [NNFI] = 1; goodness of fit index [GFI] = .98; adjusted goodness-of-fit index [AGFI] = .97) and comprises a second-order methodological quality dimension ($\omega$ = .87; $D$ = .55) that contains two first-order dimensions: quality of design (6 items; $\omega$ = .90; $D$ = .46; intraclass correlation coefficient [ICC] = .933–.967) and quality of measurement and analysis (5 items; $\omega$ = .68; $D$ = .67; ICC = .797–.988). Each item receives a score for methodological quality levels from 0 (lowest methodological quality) to 1 (highest methodological quality).

For the comparison with the MQSOM, three instruments were selected: (a) the Methodological Rigor in Mixed Methods (MRMM) checklist (Harrison et al., 2020), the GREOM (Portell et al., 2015), and the Mixed Methods Appraisal Tool (MMAT) checklist (Hong et al., 2018). Table 1 presents an overview of the instruments chosen for the comparison.

### *Procedure*

A total of nine instruments were considered for comparison to the MQSOM. The selected instruments are either guidelines or checklists from both the social and health sciences. The selection criteria were (a) type of studies the instruments evaluate (tools that assessed studies based on observational methodology and mixed-method studies were considered); (b) whether the instrument addresses the quality of reporting, design, measurement, and analysis; and (c) reliability and/or validity evidence reported. Table 2 presents the characteristics of the instruments considered.

Based on the established criteria, the GREOM was chosen because it was the only tool specific to studies based on observational methodology. Additionally, the MRMM and the MMAT were selected given that they were unique tools that assessed the quality of design, measurement, and analysis.

For the data extraction, Daniel López-Arenas (DLA) and José Mena-Raposo (JMR) were trained to apply the MQSOM, MRMM, GREOM, and MMAT. First, each item and its response options were explained. Salvador Chacón-Moscoso (SCM) mediated if the explanations differed. Then, both coders applied the instruments independently to two papers presenting studies based on observational methodology. Thereafter, DLA and JMR reviewed the coding together with SCM, who intervened in the case of discrepancies. Once the training was complete, DLA and JMR were assigned the same 24 randomly selected papers (25%) to code in order to evaluate the intercoder reliability. After achieving an adequate level of agreement (>.7), DLA applied the instruments to the full sample. Finally, 9 months after the initial coding, DLA re-coded 24 randomly selected papers (25%) to assess the intracoder reliability. The data extraction database is available as Supplementary Material.

### *Data Analysis*

Using Jeffreys's Amazing Statistics Program (JASP) 0.19.3, the ICC was calculated to evaluate and compare both inter-intracoder concordance of the instruments. Values exceeding .70 were considered adequate (Portney & Walkins, 2000). Additionally, descriptive statistics were calculated for each dimension. Subsequently, McDonald's omega ($\omega$) derived through principal factor analysis and Cronbach's alpha ($\alpha$) were both employed to assess the reliability based on the internal consistency of each instrument. Results higher than .80 were considered robust reliability evidence, and .65–.80, acceptable (Kalkbrenner, 2023). For the item discrimination, the average item discrimination index was computed. Results were considered excellent for values higher than .40, adequate for .20–.30, and inadequate for <.20 (Holgado-Tello, 2015).

Finally, a multitrait-multimethod (MTMM) analysis was conducted to provide evidence of the convergent and discriminant validity of the MQSOM (Campbell & Fiske, 1959). This approach is based on the idea that a measurement scale represents the combination of a specific trait and a particular measurement method. The correlation between scores obtained from two measurement instruments may therefore arise from two sources: (1) the fact that they assess the same underlying construct and (2) the similarity of the measurement methods used. To minimize the influence of the method effect, the MTMM approach involves assessing two or more distinct traits using two or more different measurement methods. When the same trait is measured by different methods, strong correlations are expected, providing evidence of convergent validity. Conversely, if different traits show unexpectedly high correlations, this may indicate that methodological similarities are introducing bias, thereby threatening validity. In contrast, when distinct traits are compared, only weak correlations are expected (these constitute discriminant validity evidence). Therefore, Spearman's correlation ($\rho$) matrix was constructed with the dimensions from the MQSOM and those from

[1]The list of articles included is available at https://osf.io/3zbqy/files/osfstorage.

https://doi.org/10.1017/SJP.2026.10021 Published online by Cambridge University Press

**Table 1.** Overview of the instruments for comparison

| Instrument | MRMM | GREOM | MMAT |
|---|---|---|---|
| Reference | Harrison et al. (2020) | Portell et al. (2015) | Hong et al. (2018) |
| Objective | This checklist outlines a framework for conducting mixed-methods research | This is a set of proposed guidelines for reporting studies based on observational methodology | This checklist assesses the methodological quality of mixed-methods studies, qualitative studies, randomized control trials, nonrandomized studies, and quantitative descriptive studies |
| Structure | 6 items: data collection, data analysis of qualitative and quantitative data strands, integration of both data strands, use of a specific mixed-methods design, providing a mixed-methods discussion, and elements of writing | 14 items in six dimensions: intervention and expected outcomes (2 items), design (1 item), sample (3 items), instruments (3 items), data quality control (3 items), and results (2 items) | In this study, 15 items from 3 dimensions are considered: a qualitative component (5 items), a quantitative component (5 items), and a mixed-methods component (5 items) |
| Reliability | — | — | — |
| Validity | — | — | Presents validity evidence based on the content of the items (Hong et al., 2019) |
| Relevance indicators | The original article has 138 citations in Scopus and a FWCI value of 11.5 | The original article has 82 citations in Scopus and a FWCI value of 2.1 | The original article has 1,943 citations in Scopus and a FWCI value of 43.97 |

*Note.* FWCI = field-weighted citation impact.

**Table 2.** Characteristics of the instruments considered for comparison to the MQSOM

| Instrument | Author | Year | Area | OMS | MMS | QR | QD | QM | QA |
|---|---|---|---|---|---|---|---|---|---|
| EGPQRS | Elliot et al. | 1999 | Psychology | | | X | | | |
| GRMS | Stone and Shiffman | 2002 | Medicine | | | X | | | |
| GRAMMS | O'Cathain et al. | 2008 | Public health | | X | X | | | |
| GCRMR | Leech and Onwuegbuzie | 2010 | Counseling | | X | X | | | |
| STROBE | Von Elm et al. | 2007 | Epidemiology | | | X | | | |
| COREQ | Booth et al. | 2014 | Public health | | | X | | | |
| SRQR | O'Brien et al. | 2014 | Medicine | | | X | | | |
| GREOM | Portell et al. | 2015 | Behavioral sciences | X | | X | | | |
| MRMM | Harrison et al. | 2020 | Management | | X | | X | X | X |
| MMAT | Hong et al. | 2018 | Public health | | X | | X | X | X |

*Note.* EGPQRS = Evolving Guidelines for Publication of Qualitative Research Studies in Psychology and Related Fields; GRMS = Guidelines for Reporting Momentary Studies; STROBE = Strengthening the Reporting of Observational Studies in Epidemiology; COREQ = COnsolidated criteria for REporting Qualitative Research; GRAMMS = Good Reporting of a Mixed Methods Study; GCRMR = Guidelines for Conducting and Reporting Mixed Research in the Field of Counselling and Beyond; GREOM = Guidelines for Reporting Evaluations Based on Observational Methodology; SRQR = Standards for Reporting Qualitative Research; MRMM = Methodological Rigor in Mixed Methods; OMS = suitability for assessing studies based on observational methodology; MMS = suitability for assessing mixed-methods studies; QR = suitability for assessing quality of reporting; QD = suitability for assessing quality of design; QM = suitability for assessing quality of measurement; QA = suitability for assessing quality of analysis.

the instruments for comparison included in the analysis using Spearman's correlation coefficient (ρ) due to the non-normal distribution of the scores. Correlation values over .4 were considered high, over .25 were considered moderate, and over .1 were considered low (Cohen, 1988). For convergence validity, high correlations were expected between dimensions that address similar constructs; conversely, low correlations between dimensions that address different constructs were interpreted as evidence of discriminant validity (Campbell & Fiske, 1959).

## Results

### *Inter- and Intracoder Reliability and Descriptive Statistics of the Items*

Table 3 presents both inter- and intracoder reliability and descriptive statistics. ICC coefficients were adequate, ranging from .7 to 1. Regarding the descriptive statistics, the median in MQSOM was .67 for the design dimension and .7 for the measurement and analysis dimension; the means were .62 for design and .65 for the measurement and analysis; and the standard deviations were 0.28 for design and 0.18 for measurement and analysis. Regarding MRMM, the median was 1 or .5 for all the dimensions, with 0 being the median only for the dimension of mixed-methods design type. The means ranged from .43 to .90, and the standard deviations were between 0.17 and 0.46.

Regarding the GREOM, the median was .5 or higher for 66.7% of the dimensions, with a 0 only for design. The means were between .43 and .68, and the standard deviations ranged between 0.21 and 0.46. Regarding the MMAT, the median was .6 for the QUAL and mixed-methods component and .7 for the QUANT component. The means were between .56 and .65, and the standard deviations ranged between .18 and .23. Normal distribution was not found for any of the items of the different instruments.

https://doi.org/10.1017/SJP.2026.10021 Published online by Cambridge University Press

**Table 3.** Descriptive statistics and inter–intracoder reliability of the instruments for comparison

| | Intercoder reliability | | | Intracoder reliability | | | Descriptive statistics | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dimension | Avg. | Min. | Max. | Avg. | Min. | Max. | *M* | *Mdn* | *SD* | S | K | KS |
| MQSOM | | | | | | | | | | | | |
| Design | .950 | .933 | .967 | .725 | .725 | .976 | .62 | .67 | 0.28 | −0.35 | −1.20 | .157 |
| M&A | .936 | .797 | .988 | .903 | .858 | .963 | .65 | .70 | 0.18 | −0.45 | 0.15 | .216 |
| MRMM | | | | | | | | | | | | |
| Aims and purpose[a] | .808 | — | — | .753 | — | — | .84 | 1 | 0.23 | −0.82 | −1.35 | .436 |
| Data collection[a] | .780 | — | — | .775 | — | — | .54 | .5 | 0.17 | 1.37 | 5.46 | .493 |
| Data analysis[a] | .733 | — | — | .880 | — | — | .90 | 1 | 0.20 | −1.54 | 0.38 | .491 |
| Data integration[a] | .763 | — | — | .859 | — | — | .52 | .5 | 0.24 | 0.12 | 1.48 | .399 |
| MM design[a] | .939 | — | — | .987 | — | — | .43 | 0 | 0.46 | 0.30 | −1.80 | .332 |
| Writing[a] | .903 | — | — | .884 | — | — | .75 | 1 | 0.36 | −1.06 | −0.29 | .384 |
| GREOM | | | | | | | | | | | | |
| Intervention | .845 | .727 | .962 | .734 | .730 | .737 | .64 | .75 | 0.23 | −0.23 | −0.77 | .254 |
| Design[a] | .880 | — | — | 1 | — | — | .43 | 0 | 0.46 | 0.30 | −1.80 | .332 |
| Samples | .844 | .718 | .940 | .804 | .720 | .851 | .52 | .5 | 0.21 | 0.01 | −1.04 | .192 |
| Instruments | .821 | .704 | .912 | .865 | .754 | 1 | .68 | .67 | 0.26 | −0.40 | −0.86 | .182 |
| Data quality | .881 | .762 | .984 | .864 | .728 | 1 | .47 | .33 | 0.28 | 0.27 | −0.42 | .260 |
| Results | .825 | .720 | .930 | .785 | .775 | .795 | .54 | .50 | 0.25 | 0.38 | 0.81 | .392 |
| MMAT | | | | | | | | | | | | |
| Qualitative | .800 | .707 | .880 | .898 | .726 | 1 | .62 | .60 | 0.23 | −0.04 | −0.40 | .115 |
| Quantitative | .863 | .779 | .943 | .955 | .857 | 1 | .65 | .70 | 0.21 | −1.01 | 1.20 | .200 |
| Mixed methods | .818 | .726 | .896 | .971 | .854 | 1 | .56 | .60 | 0.18 | −0.16 | −0.69 | .202 |

*Note.* Avg. = average of intraclass correlation coefficients; Min. = minimum; Max. = maximum; *M* = mean; Mdn = median; *SD* = standard deviation; S = skewness; K = kurtosis; KS = Kolmogorov–Smirnov normality test; MQSOM design = quality of design; MQSOM M&A = quality of measurement and analysis; MRMM MM design = mixed-methods design type; MRMM writing = elements of writing; GREOM intervention = intervention and expected outcomes.
[a]ICC is computed for a single item, and there is no minimum or maximum.
ICC and KS obtained $p < .05$.

### *Reliability Based on Internal Consistency and Discrimination*

Table 4 presents the results obtained for the reliability coefficients based on the internal consistency and discrimination indexes of each dimension of the instruments under consideration.

The reliability based on the internal consistency of the instruments was adequate only for the MQSOM's design and the MMAT QUAL and mixed-methods components. Discrimination was considered excellent for all dimensions except MQSOM's measurement and analysis, and the MRMM's aims and purpose and data integration, which were considered adequate, and the MRMM's data collection and data analysis, which were deemed poor.

### *MTMM Matrix*

Table 5 presents the MTMM matrix with Spearman's correlation coefficients of the MQSOM dimensions with the dimensions of MRMM, GREOM, and MMAT.

The results showed significant correlations of MQSOM with 63.33% of the instrument dimensions (19 correlations), with magnitudes ranging from .21 to .64. Additionally, a significant moderate factor correlation was identified within MQSOM ($\rho = .364$, $p < .001$).

Regarding the MRMM checklist, the MQSOM dimension of design exhibited moderate correlations with the dimensions of aims and purpose ($\rho = .23$, $p = .027$) and mixed-methods design type ($\rho = .22$, $p = .034$) and a strong correlation with elements of writing ($\rho = .47$, $p < .001$). Conversely, it showed low and nonsignificant correlations with the dimensions of data collection ($\rho = .03$, $p = .807$) and data analysis ($\rho = -.05$, $p = .620$). Regarding the GREOM, the MQSOM design dimension exhibited a moderate correlation with the design dimension ($\rho = .22$, $p = .034$) and a moderate-to-strong correlation with the intervention and expected outcome dimension ($\rho = .34$, $p = .001$). Conversely, it exhibited low and nonsignificant correlations with samples ($\rho = -.03$, $p = .750$), data quality control ($\rho = .03$, $p = .812$), and results ($\rho = -.03$, $p = .759$). Regarding the MMAT checklist, the MQSOM dimension design exhibited a moderate correlation with the QUAL component ($\rho = .21$, $p = .038$). Conversely, it exhibited a low and nonsignificant correlation with the QUANT component ($\rho = -.04$, $p = .733$).

Regarding the MRMM checklist, the MQSOM dimension measurement and analysis exhibited a moderate correlation with the aims and purpose dimension ($\rho = .21$, $p = .041$) and strong correlations with the dimensions data integration ($\rho = .48$, $p < .001$), mixed-methods design type ($\rho = .45$, $p < .001$), elements of writing ($\rho = .42$, $p < .001$), and data analysis ($\rho = .61$, $p < .001$). Conversely, it showed low-to-moderate correlation and nonsignificant relationships with the dimensions data collection ($\rho = .17$, $p = .105$).

https://doi.org/10.1017/SJP.2026.10021 Published online by Cambridge University Press

**Table 4.** Reliability and discrimination of the dimensions

| Item | McDonald's ω | 95% CI | | Cronbach's α | 95% CI | | Discrimination | | |
|---|---|---|---|---|---|---|---|---|---|
| | | LL | UL | | LL | UL | Avg. | Min. | Max. |
| MQSOM | | | | | | | | | |
| Design | .818 | .770 | .857 | .741 | .652 | .811 | .66 | .19 | 1 |
| M&A | .342 | .023 | .759 | .271 | .006 | .476 | .33 | .12 | .50 |
| MRMM | | | | | | | | | |
| Aims and purpose[a] | — | — | — | — | — | — | .37[b] | — | — |
| Data collection[a] | — | — | — | — | — | — | .13[b] | — | — |
| Data analysis[a] | — | — | — | — | — | — | .25[b] | — | — |
| Data integration[a] | — | — | — | — | — | — | .38[b] | — | — |
| MM design[a] | — | — | — | — | — | — | .94[b] | — | — |
| Writing[a] | — | — | — | — | — | — | .62[b] | — | — |
| GREOM | | | | | | | | | |
| Intervention | .281 | .051 | .439 | .273 | −.039 | .502 | .53 | .38 | .67 |
| Design[a] | — | — | — | — | — | — | 1[b] | — | — |
| Samples | .068 | .001 | .358 | .056 | −.280 | .319 | .50 | .23 | 1 |
| Instruments | .524 | .369 | .749 | .498 | .312 | .641 | .64 | .38 | .92 |
| Data quality | .360 | .168 | .580 | .345 | .082 | .542 | .61 | .44 | .73 |
| Results | .265 | .161 | .356 | .261 | −.072 | .501 | .48 | .35 | .62 |
| MMAT | | | | | | | | | |
| Qualitative | .754 | .707 | .967 | .734 | .647 | .804 | .54 | .29 | .96 |
| Quantitative | .585 | .447 | .909 | .468 | .267 | .622 | .48 | .23 | .71 |
| Mixed methods | .675 | .521 | .747 | .647 | .529 | .741 | .45 | .10 | .67 |

*Note.* CI = confidence interval; LL = lower limit; UL = upper limit; Avg. = average of intraclass correlation coefficients; Min. = minimum; Max. = maximum; MQSOM Design = quality of design; MQSOM M&A = quality of measurement and analysis; MRMM MM design = mixed-methods design type; MRMM writing = elements of writing; GREOM intervention = intervention and expected outcomes.
[a]Reliability coefficients are not computed since the dimension is comprised of a single item.
[b]The discrimination indexes displayed are not an average but unitary since they are formed by a single item.

Regarding the GREOM, the measurement and analysis dimension exhibited strong correlations with the dimensions design (ρ = .45, $p < .001$), data quality control (ρ = .42, $p < .001$), results (ρ = .33, $p = .001$), intervention and expected outcomes (ρ = .55, $p < .001$), and instruments (ρ = .61, $p < .001$). Conversely, it showed a low and nonsignificant correlation with the dimension samples (ρ = .11, $p = .276$). Regarding the MMAT checklist, the measurement and analysis dimension exhibited a moderate correlation with the QUANT component (ρ = .21, $p = .037$) and strong correlations with both the QUAL (ρ = .64, $p < .001$) and mixed-methods (ρ = .64, $p < .001$) components.

## Discussion

The findings provide evidence of adequate inter- and intracoder reliability for the MQSOM and the instruments for comparison. In contrast, both McDonald's omega and Cronbach's alpha values indicate a lack of internal consistency for the MQSOM's measurement and analysis dimension, all the GREOM dimensions, and the MMAT's QUANT component. Regarding discrimination, the global average discrimination index was deemed acceptable for the MQSOM and all the dimensions of the instruments for comparison, except for the MRMM's data collection and data analysis. This suggests that these dimensions (which comprise a single item each) may be ineffective in distinguishing between high- and low-quality studies.

The MTMM analysis provides substantial evidence of the convergent and discriminant validity of MQSOM. Specifically, the MQSOM dimensions demonstrated significant correlations with dimensions addressing similar constructs through all instruments for comparison, ranging from moderate to strong. Conversely, the design quality dimension demonstrated low, nonsignificant correlations with the dimensions that assess different constructs (Campbell & Fiske, 1959).

Specifically, the MQSOM's design demonstrates convergent validity evidence with the same constructs from the instrument for comparison, e.g., the MRMM's and GREOM's designs. Additionally, it demonstrates convergent validity evidence with several analogous constructs. These include MRMM's aims and purpose dimension, which involves justifying the use of mixed methods, and its elements of writing dimension, which encompass identifying the study as mixed methods. The MQSOM's design also shows convergence with GREOM's intervention and expected outcomes, which requires justification for employing an observational methodology, and with MMAT's QUAL component. This latter convergence is consistent with the latent construct of design quality addressed by MQSOM, since the initial QUAL phase begins with a justification for the use of observational methodology (Anguera et al., 2020).

https://doi.org/10.1017/SJP.2026.10021 Published online by Cambridge University Press

**Table 5.** Multitrait-multimethod matrix

| | MQSOM design | MQSOM measurement and analysis |
|---|---|---|
| MQSOM measurement and analysis | .364** | |
| MRMM1 aims and purpose | .225* | .209* |
| MRMM2 data collection | .025 | .167 |
| MRMM3 data analysis | −.051 | .611** |
| MRMM4 data integration | −.096 | .482** |
| MRMM5 MM design type | .217* | .445** |
| MRMM6 elements of writing | .468** | .419** |
| GREOM1 intervention and expected outcomes | .339** | .553** |
| GREOM2 design | .217* | .445** |
| GREOM3 samples | −.033 | .112 |
| GREOM4 instruments | .196 | .607** |
| GREOM5 data quality control | .025 | .423** |
| GREOM6 results | −.032 | .328** |
| MMAT QUAL component | .212* | .636** |
| MMAT QUANT component | −.035 | .214* |
| MMAT mixed-methods component | .190 | .643** |

*Note.* MQSOM design = quality of design; MRMM5 MM design type = mixed-methods design type.
*$p < .05$; **$p < .01$.

The MQSOM's measurement and analysis demonstrate convergent validity evidence with the same constructs from the instrument for comparison, such as MRMM's data analysis and GREOM's instruments, data quality control, and results. Additionally, convergent validity evidence was exhibited with similar constructs, such as MRMM's data integration and MMAT's QUAL, QUANT, and mixed-methods components. This finding is congruent with the latent construct quality of measurement and analysis assessed by the MQSOM. In the context of observational methodology, measurement entails the coding of spontaneous behavior to yield a matrix of QUAL data; the integration of data strands is achieved when these QUAL data are transformed into QUANT parameters for analysis, positioning observational methodology as a mixed-methods approach (Anguera et al., 2020).

Regarding discriminant validity evidence, the MQSOM's design obtained low and nonsignificant correlations with MRMM's data collection, data integration, and data analysis, GREOM's samples, instruments, data quality control, and results, and the MMAT's QUANT and mixed-methods components. Additionally, the MQSOM's measurement and analysis obtained discriminant evidence with GREOM's samples.

On the other hand, MQSOM's measurement and analysis yielded some unexpected results. First, significant correlations were exhibited with MRMM's both aims and purpose (although low–moderate) and elements of writing (strong). This can be attributed to the pivotal role of the discussion section in the observational methodology process, given that the final step is the QUAL interpretation of QUANT results. The discussion section is also critical in the coding criteria of the MRMM items, which assess the value of mixed-method research for aims and purpose, and in the incorporation of references to mixed-method studies for elements of writing.

Additionally, the MQSOM's measurement and analysis dimension exhibited significant, strong correlations with the design dimensions of both MRMM and GREOM. This can be attributed to the inherent overlap between research design and research implementation, whereby the quality of the design can influence the quality of subsequent measurement and analysis. Finally, there was a nonsignificant correlation between the MQSOM's measurement and analysis and the analogous construct, the MRMM's data collection. This could be attributed to the fact that the MQSOM is centered on methodological quality (e.g., the adequacy of the instrument), while the MRMM focuses on the quality of the reporting (e.g., if the instrument used is specified).

### *Study Limitations and Further Development*

Although this study has contributed to the validity evidence of the MQSOM, several limitations should be considered. First, the internal consistency of MQSOM's measurement and analysis was below optimal levels. This finding, unexpected based on the results obtained in the original validation of the scale (Sanduvete-Chaves et al., 2025) and based on other applications with a larger sample, indicates a potential need for further research with a larger and more diverse sample of studies based on observational methodology to ensure the reliable measurement of the latent construct. In the present study, we decided to maintain the validity of the dimension in its original version instead of introducing modifications to obtain a higher reliability coefficient, given that the dimension is considered well represented; e.g., each item refers to a different aspect: observation instrument, software, type of parameter, reliability, and type of data analysis.

Additionally, the instruments available for comparison in the mixed methods, such as guidelines and checklists, do not always provide comprehensive psychometric evidence, particularly regarding reliability and construct validity. Furthermore, they focus mainly on the quality of the reporting, while the MQSOM is mainly centered on measuring methodological quality. Consequently, while the evidence supports convergent and discriminant validity evidence, these findings should be interpreted within the context of the differences in the constructs that the instruments compared are measuring. Further research is necessary to compare the MQSOM with other scales that may be developed in the coming years to assess the methodological quality in mixed-methods studies.

Furthermore, the preliminary evidence of convergent and discriminant validity based on correlation analyses should be complemented with additional statistical techniques—such as confirmatory factor analysis, structural equation modeling, or network analysis—to provide a deeper understanding of the latent dimensions in which the MQSOM and the comparison instrument appear (or fail) to converge (Marsh & Grayson, 1995). Future studies should also examine the factor invariance of the MQSOM across diverse areas of applied research grounded in observational methodology.

From a practical standpoint, the MQSOM scale can be applied across a wide range of research contexts, such as studies in sports—including soccer (López-Arenas et al., in press), tennis, and basketball. These applications will contribute to synthesizing the accumulated empirical evidence from various lines of research based on observational methodology. Consequently, distinct methodological quality profiles will be available for each research area, facilitating the comparison and extrapolation of high-quality observational research across disciplines.

https://doi.org/10.1017/SJP.2026.10021 Published online by Cambridge University Press

## Conclusions

In summary, this study contributes to an understanding of the psychometric properties of the MQSOM, providing convergent and discriminant validity evidence. The findings indicate high and significant correlations with equal or similar constructs measured with the MQSOM and the instruments for comparison and low and nonsignificant correlations in relation to different constructs. This reinforces the validity evidence of the MQSOM as a useful methodological quality scale.

**Supplementary material.** The supplementary material for this article can be found at http://doi.org/10.1017/SJP.2026.10021.

**Data availability statement.** The datasets presented in this study can be found as supplementary material and in the online repository Open Science Framework (https://osf.io/3zbqy/).

**Acknowledgements.** We would like to thank Wendy Gosselin (American Translation Association member #275293) for reviewing the English version of this text for publication.

**Author contribution.** S.C.M. and S.S.C conceptualized the study. D.L.A. and J.M.R. investigated the data. D.L.A. and J.M.R. curated the data. D.L.A. is involved in formal analysis. D.L.A. wrote the original draft. S.C.M. and S.S.C. supervised the data. S.S.C. and S.C.M. wrote, reviewed, and edited the manuscript.

**Funding statement.** This work was supported by ANID, Government of Chile, the National Scientific and Technological Development Fund (FONDECYT) project 1250316 (S.C.M. and S.S.C.); the Julio Olea Grant for Young Researchers of the Spanish Association of Behavioral Science Methodology (AEMCCO; D.L.A.); and the Spanish Ministry of Science, Innovation and Universities (MICIU/AEI/10.13039/501100011033), research project PID2020-115486GB-I00 (all authors). Funding for open access charge: Universidad de Sevilla/CBUA.

**Competing interests.** None declared.

## References

**Anguera, M. T.**, **Blanco-Villaseñor, A.**, **Losada, J. L.**, & **Portell, M.** (2018). Guidelines for designing and conducting a study that applies observational methodology. *Anuario de Psicología*, **48**(1), 9–17. https://doi.org/10.1016/j.anpsic.2018.02.001

**Anguera, M. T.**, **Blanco-Villaseñor, A.**, **Losada, J. L.**, & **Sánchez-Algarra, P.** (2020). Integración de elementos cualitativos y cuantitativos en metodología observacional [Integration of qualitative and quantitative elements in observational methodology]. *Ámbitos: Revista Internacional de Comunicación*, **49**, 49–70. https://doi.org/10.12795/Ambitos.2020.i49.04

**Anguera, M. T.**, **Camerino, O.**, **Castañer, M.**, **Sánchez-Algarra, P.**, & **Onwuegbuzie, A.** (2017). The specificity of observational studies in physical activity and sports sciences: Moving forward in mixed methods research and proposals for achieving quantitative and qualitative symmetry. *Frontiers in Psychology*, **9**(2196). https://doi.org/10.3389/fpsyg.2017.02196

**Booth, A.**, **Hannes, K.**, **Harden, A.**, **Noyes, J.**, **Harris, J.**, & **Tong, A.** (2014). Chapter 21. COREQ (consolidated criteria for reporting qualitative studies). In D. Moher, D. G. Altman, K. F. Schulz, I. Simera, & E. Wager (Eds.), *Guidelines for reporting health research: A user's manual* (pp. 214–226). John Wiley & Sons.

**Campbell, D. T.**, & **Fiske, D. W.** (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, **56**(2), 81–105. https://doi.org/10.1037/h0046016

**Chacón-Moscoso, S.**, **Anguera, M. T.**, **Sanduvete-Chaves, S.**, **Losada, J. L.**, & **Portell, M.** (2019). Methodological quality checklist for studies based on observational methodology (MQCOM). *Psicothema*, **31**(4), 458–464. https://doi.org/10.7334/psicothema2019.116

**Chacón-Moscoso, S.**, **Sanduvete-Chaves, S.**, **Portell, M.**, & **Anguera, M. T.** (2013). Reporting a program evaluation: Needs, program plan, intervention, and decisions. *International Journal of Clinical and Health Psychology*, **13**(1), 58–66. https://doi.org/10.1016/S1697-2600(13)70008-5

**Cochran, W. G.**, & **Chambers, S. P.** (1965). The planning of observational studies of human populations. *Journal of the Royal Statistical Society. Series A (General)*, **128**(2), 234–266. https://doi.org/10.2307/2344179

**Cohen, J.** (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates.

**Cresswell, J. W.**, & **Plano-Clark, V.** (2017). *Designing and conducting mixed methods research*. Sage.

**Elliott, R.**, **Fischer, C. T.**, & **Rennie, D. L.** (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*, **38**, 215–229. https://doi.org/10.1348/014466599162782

**Harrison, R.**, **Reilly, T.**, & **Creswell, J.** (2020). Methodological rigor in mixed methods: An application in management studies. *Journal of Mixed Methods Research*, **14**(4), 473–495. https://doi.org/10.1177/1558689819900585

**Holgado-Tello, F. P.** (2015). Análisis de los ítems. [analysis of the items]. In M. I. Barbero-García, E. Vila-Abad, & F. P. Holgado-Tello (Eds.), *Psicometría* [Psychometrics] (pp. 407–468). Sanz y Torres.

**Hong, Q. N.**, **Fàbregues, S.**, **Bartlett, G.**, **Boardman, F.**, **Cargo, M.**, **Dagenais, P.**, **Gagnon, M. P.**, **Griffiths, F.**, **Nicolau, B.**, **O'Cathain, A.**, **Rousseau, M. C.**, & **Vedel, I.** (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, **34**(4), 285–291. https://doi.org/10.3233/EFI-180221

**Hong, Q. N.**, **Pluye, P.**, **Fàbregues, S.**, **Bartlett, G.**, **Boardman, F.**, **Cargo, M.**, **Dagenais, P.**, **Gagnon, M. P.**, **Griffiths, F.**, **Nicolau, B.**, **O'Cathain, A.**, **Rousseau, M. C.**, & **Vedel, I.** (2019). Improving the content validity of the mixed methods appraisal tool: A modified e-Delphi study. *Journal of Clinical Epidemiology*, **111**, 49–59. https://doi.org/10.1016/j.jclinepi.2019.03.008

**Kalkbrenner, M. T.** (2023). Alpha, omega, and H internal consistency reliability estimates: Reviewing these options and when to use them. *Counseling Outcome Research and Evaluation*, **14**(1), 77–88. https://doi.org/10.1080/21501378.2021.1940118

**Leech, N.**, & **Onwuegbuzie, A.** (2010). Guidelines for conducting and reporting mixed research in the field of counseling and beyond. *Journal of Counseling & Development*, **88**(1), 61–69. https://doi.org/10.1002/j.1556-6678.2010.tb00151.x

**Marsh, H. W.**, & **Grayson, D.** (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and application* (pp. 177–198). Sage.

**O'Brien, B. C.**, **Harris, I. B.**, **Beckman, T. J.**, **Reed, D. A.**, & **Cook, D. A.** (2014). Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine*, **89**, 1245–1251. https://doi.org/10.1097/ACM.0000000000000388

**O'Cathain, A.**, **Murphy, E.**, & **Nicholl, J.** (2008). The quality of mixed methods studies in health services research. *Journal of Health Services Research & Policy*, **13**(2), 92–98. https://doi.org/10.1258/jhsrp.2007.007074

**Portell, M.**, **Anguera, M. T.**, **Chacón-Moscoso, S.**, & **Sanduvete-Chaves, S.** (2015). Guidelines for reporting evaluations based on observational methodology. *Psicothema*, **27**(3), 283–289. https://doi.org/10.7334/psicothema2014.276

**Portney, L.**, & **Watkins, M.** (2000). *Foundations of clinical research: Applications to practice*. Prentice Hall.

**Sanduvete-Chaves, S.**, **López-Arenas, D.**, **Anguera, M. T.**, & **Chacón-Moscoso, S.** (2025). A scale for evaluating the methodological quality of studies based on observational methodology. *Psicothema*, **37**(1), 1–10. https://doi.org/10.70478/psicothema.2025.37.01

**Stone, A. A.**, & **Shiffman, S.** (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, **147**(8), 573–578. https://doi.org/10.1207/S15324796ABM2403_09

**von Elm, E.**, **Altman, D. G.**, **Effer, M.**, **Pocock, S. J.**, **Gotzsche, P. C.**, & **Vandenbroucke, J. P.** (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Annals of Internal Medicine*, **147**(8), 573–577. https://doi.org/10.7326/0003-4819-147-8-200710160-00010