Test-retest reliability and criterion validity of the Spanish Version of two Motor Imagery questionnaires in people with Parkinson Disease

Abstract

Background and purpose. The Kinesthetic and Visual Imagery Questionnaire (KVIQ) and Movement Imagery Questionnaire-Revised Second version (MIQ-RS) are measurement instruments that assess motor imagery vividness. The aim was to examine the validity and reliability of the Spanish KVIQ and MIQ-RS in people with Parkinson Disease (PD).

Methods. A longitudinal descriptive study was conducted following the COSMIN standards. Thirty-five people with idiopathic PD were assessed twice (7-15 days apart) with the Spanish KVIQ and MIQ-RS. Structural validity, internal consistency, test-retest reliability (ICC), standard error of measurement (SEM), smallest detectable change (SDC) and criterion validity of the MIQ-RS and KVIQ long (KVIQ-20), short (KVIQ-10) and extended (KVIQ-34) versions and their subscales (if pertinent) were tested.

Results. Factor Analysis was satisfactory for the MIQ-RS, KVIQ-20 and KVIQ-10 questionnaires, providing evidence of their two-dimensional structure. Evidence of the structural validity of the KVIQ-34 was not confirmed and thus was analysed as an overall score. Revelle's Omega >0.9 showed excellent internal consistency. Test-retest reliability was moderate (ICC = 0.58-0.75) and higher for all visual subscales. SEM and

SDC were up to 14.39% and 39.89% of the scores, respectively. Criterion validity between questionnaires and subscales was strong (Spearman's r >0.7).

Discussion and conclusions. The results provide evidence of the validity and reliability of the Spanish MIQ-RS, KVIQ-20 and KVIQ-10 to assess motor imagery vividness in people with PD, whereas the KVIQ-34 should only be interpreted as an overall score. Psychometric, procedural and practical features of the questionnaires should be considered when applying them into clinical practice.

Key words: Motor Imagery; Vividness; Assessment; Parkinson Disease; Test-retest reliability

The Video Abstract of this article can be consulted at this URL.

Introduction

Mental practice with Motor Imagery (MI) has been widely and successfully used to improve motor performance in both sports[1,2] and musical disciplines[3] through the promotion of motor learning.[4] However, evidence of its effectiveness on neurodegenerative diseases such as Parkinson Disease (PD) is conflicting. Several studies have shown promising results, with body schema,[5] mobility, motor function[6] or gait speed[7] improvements, whereas others have found no significant differences in comparison to relaxation[8] or conventional physiotherapy[9] interventions. The ability of the person to perform MI has been identified as one of the possible factors that may influence the effectiveness of MI-based programmes.[10] Hence, assessing MI ability is relevant in order to efficiently implement these procedures in specifically suitable clinical groups which might benefit from them to a greater extent.

MI ability can be assessed with biological or behavioural procedures. Behavioural assessments evaluate specific domains of motor images such as temporal accuracy (i.e., the time congruence between actual and imaged actions), controllability (i.e., the ability to mentally manipulate images) or vividness (i.e., the subjective intensity with which a person experiences MI). These aspects are assessed with separate tests such as mental chronometry, mental rotation or questionnaires, respectively.[11,12] Specifically, vividness questionnaires are widespread tools typically distinguishing between the two main sensory modalities used during imagery (i.e., visual and kinesthetic).[13] There are plenty of questionnaires currently available,[14] but two of the most widely used in clinical contexts are the Kinesthetic and Visual Imagery Questionnaire (KVIQ) and the Movement Imagery Questionnaire-Revised Second Version (MIQ-RS), both specifically developed for people with sensory-motor disabilities.[15,16] Evidence of the validity of the English KVIQ in people with PD is available,[17] whereas there is a lack of

studies providing evidence of the validity and reliability of the MIQ-RS. These tools represent an inexpensive, fast and useful procedure to detect poor or good imagers and therefore may optimise MI-based treatments by identifying who is more suitable to be trained with mental practice techniques and selecting therapeutic strategies based on performance.

Multiple trans-cultural adaptations and translations into several languages are now available for both questionnaires,[18–21] as well as evidence of good psychometric properties in people with stroke or Multiple Sclerosis.[15,22,23] Although the KVIQ has been used in people with PD,[24–27] the usage of the MIQ-RS is much more limited due to a paucity of psychometric studies. Previous work consistently shows that both questionnaires have a two-dimensional structure, therefore assessing visual or kinesthetic vividness separately.[15,16,23] There is also evidence of excellent internal consistency (Cronbach's alpha = 0.87-0.98), acceptable to moderate test-retest reliability (ICC: KVIQ = 0.81-0.9; MIQ-RS = 0.54-0.73) and good criterion validity between them.[15,23] Nonetheless, their psychometric behaviour remains largely unexamined in people with PD. No studies to date have analysed the validity of the MIQ-RS in these people, though Randhawa, Harris & Boyd (2007) assessed test-retest reliability of the English KVIQ and its criterion validity with the MIQ-R (first revised version) in a small sample of mild to moderate PD participants (n = 11).[17] However, they did not evaluate its factor structure or internal consistency which should be previously examined according to current consensus-based standards.[28] In addition, the MIQ-R includes difficult movements (e.g., jumping) and thus is not the most appropriate questionnaire in people with sensory-motor deficits, who should be assessed with the MIQ-RS.[29]

Spanish translations of the MIQ-RS and KVIQ questionnaires are recently available, which allows to examine their appropriateness to evaluate MI ability in Spanish-speaking clinical populations.[30,31] Therefore, the purpose of this study was to assess the psychometric properties of these two Spanish versions of the questionnaires in people with PD. We aimed to (1) assess structural validity to evaluate their two-dimensional structure (2) examine internal consistency and test-retest reliability of total and subscale scores individually, if pertinent (3) evaluate criterion validity by correlating their total and subscale scores.

Methods

Study design

This study followed the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines.[28] A descriptive longitudinal study at the Madrid Parkinson Association was carried out. The study was approved by the Ethical Committee for Clinical Research of the San Carlos Hospital of Madrid and written informed consent from all participants was obtained prior to enrolment. All procedures were in accordance with the 1964 Declaration of Helsinki.

Participants

Sample size calculation

Sample size was calculated using previous models proposed for test-retest reliability studies using an asymmetrical interval procedure for not achieving a prespecified lower limit.[32] The prespecified value (p_0) was ICC ≤ 0.5 for unacceptable reliability,[33] with a 95% probability of not obtaining that value. The expected value of ICC was ≥ 0.8 based on previous literature.[17] Thirty-two participants were needed to reach a significance

level of 0.05 and test power of 0.8, but considering a dropout rate of 10%, 35 participants were required. This calculation was performed in Microsoft Excel 2016.

Selection criteria

Participants were included when they were (1) diagnosed of idiopathic PD (according to the United Kingdom Parkinson Disease Society Brain Bank Criteria),[34] (2) older than 60 years and (3) able to independent standing. Participants were excluded if there was evidence of (1) cognitive impairment (MiniMental State Examination < 24)[35] or (2) diagnosis of other neurological diseases than PD, psychiatric diseases, orthopaedic or cardiovascular diseases, or (3) presence of visual sensory deficits (i.e., complete blindness) or range of motion limitations (i.e., <80% of age-adjusted active range-of-motion[36] in joints involved in movements assessed with the questionnaires, measured with standard goniometry procedures[37]) which could interfere with the assessments.

Participants' characteristics

Thirty-five people diagnosed with idiopathic PD (17 women, age = 74.2 (7.2) years), with mean disease duration of 8.8 (5.9) years and in Hoehn and Yahr stages I-IV participated. Impact of disease (MDS-UPDRS) and level of functional independence (Schwab & England scale) are shown in Table 1. According to MDS-UPDRS data,[38] participants were subdivided into Postural Instability/Gait Difficulty (PIGD) (n=25) and Tremor Dominant (TD) (n=9) motor phenotypes of PD. One participant was classified as indeterminate. None of the participants had undergone Deep Brain Stimulation surgery.

---insert Table 1 about here---

Assessments

KVIQ Spanish Version

The KVIQ questionnaire evaluates MI vividness with a difference between visual and kinesthetic sensory modalities. It can be administered as short (KVIQ-10), long (KVIQ-20), and extended (KVIQ-34) versions, which are additive.[15] It evaluates simple movements of the neck, trunk and upper and lower limbs. Limb items are administered unilaterally (short and long versions), on the dominant or non-dominant side depending on the item, or bilaterally (extended version). Each item involves four steps: 1) adopting an initial position; 2) physically performing a movement; 3) returning to the initial position, and 4) visually or kinesthetically imaging that movement. After that, the participant rates the intensity perceived during the mental task on a 5-point scale, from 1=no image to 5=image as clear as actually seeing it (visual items) and from 1=no sensation to 5=as intense as making the movement (kinesthetic items). All items are first visually imaged (visual subscale) and then kinesthetically (kinesthetic subscale). Performance on the questionnaire can be reported with total scores ranging from 10 to 50 (KVIQ-10), 20 to 100 (KVIQ-20) or 34 to 170 (KVIQ-34), subscale scores (divide by two the previous ranges) and mean individual item scores (range 1 to 5), where higher values mean better ability.

MIQ-RS Spanish Version

It is a self-administered tool that assesses imagery ease/difficulty.[16] It has one version of 14 items with two 7-item subscales (visual and kinesthetic) whose items are interspersed during the administration. Each item follows the same procedure as in the KVIQ. The participant rates the easiness or difficulty of generating the image on a 7-point scale from 1=very hard to see/feel to 7=very easy to see/feel, rather than the self-assessed intensity of the visual or kinesthetic sensations. Scores can be recorded as total

(14 to 98), subscale (7 to 49) or individual item (1 to 7) values, where higher indicate better ability.

Experimental procedure

All participants were assessed twice (7- to 15-day interval) with the paper based Spanish KVIQ and MIQ-RS questionnaires by the same experienced examiner at the Association facility. All participants were novel to MI techniques and were not explicitly instructed to perform or not to perform mental practice between assessments. Participants were not asked to change their regular medication schedule and were evaluated in the "on" medication state (i.e., one to two hours after the anti-parkinsonian medication intake).[39] Interviews confirmed that medication status did not change during the period of study participation. To avoid medication-dosing effects, all participants were tested at the same time of day for each of the two sessions and environmental conditions remained constant during both assessments. Administration of both questionnaires followed the recommendations and procedures described by their original authors. In the first assessment day the KVIQ-extended version was administered first and then the MIQ-RS questionnaire. In the second assessment day the order was counterbalanced. We excluded participants from the analyses if any data was lost after the assessments or from test to retest.

Psychometric assessment and statistical analysis

The Statistical Package for the Social Sciences (SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.) was used. All analyses were conducted with 95% confidence intervals and alpha = 0.05 for statistical significance. Shapiro-Wilk test for data normality indicated that non-parametric statistics should be used.

Assessing structural validity is an essential step as a part of the psychometric analysis according to COSMIN standards.[28] Although it is recommended that sample size should be at least five times the number of items (or n > 100) to perform Factor Analysis, given the lack of this analysis in previous studies validating MI questionnaires in people with PD,[17] we assessed the internal structure of each questionnaire and version (KVIQ) with an Exploratory Factor Analysis. Kaiser-Meyer-Olkin and Bartlett's sphericity tests were performed to determine if Factor Analysis should be used, with a cut point of >0.5 and $p<0.01$, respectively. The Principal Axis Factoring extraction method with Promax rotation was used. If the bidimensional structure of the questionnaire/version was confirmed, the next analyses were conducted for both total and subscale scores, independently.

Reliability was first tested as Internal Consistency. We used Revelle's Omega total coefficient, which is optimal when data are non-normally distributed and the questionnaire is not unidimensional nor tau-equivalent.[40,41] We used the RStudio software (RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/) via calling the "omega" function from the "psych" R package to obtain Omega coefficient.[41] Test-retest reliability was tested with the ICC using a two-way mixed-effects model with absolute agreement of single measures (3,1).[33] According to COSMIN standards, there is currently no consensus on whether ICCs should be replaced by non-parametric alternatives when non-normal distributions are present,[42] and thus we used this measure as the best consensus-based approach.[28] Acceptable values were Omega > 0.7 and ICC > 0.5.

Limits of Agreement were assessed with differences between first day and second day tests plotted against the means of the two measurements by Bland-Altman plots.[43] Measurement error was also tested using Standard Error of Measurement (SEM) and

Smallest Detectable Change (SDC) with a 95% confidence interval,[44] as SEM=SD*√1-ICC and SDC= SEM*1.96*√2.

Criterion validity between questionnaires was assessed by correlating KVIQ total and subscale scores (if pertinent) to their MIQ-RS counterparts. Criterion validity of the short version of the KVIQ was tested with the long version as the gold standard.[45] The aforementioned analyses were tested with the Spearman's rank correlation coefficient with standard interpretation[46] using the Statgraphics 19® software (Statgraphics Technologies, Inc.).

Results

Thirty-five participants initially enrolled in the study completed both assessment sessions, with a mean interval of 10.43 (2.56) days between test and retest, without any missing data or items. Participants did not report having practised MI within the study period.

Motor Imagery ability

Mean total, visual and kinesthetic scores of the different KVIQ versions (long, short and extended), as well as MIQ-RS, are listed in Table 2. According to the Wilcoxon signed-rank test there were no significant differences between visual and kinesthetic scores in any of the KVIQ versions ($p = 0.929$, 0.97 and 0.606 for the KVIQ-20, KVIQ-10 and KVIQ-34, respectively) or the MIQ-RS questionnaire ($p = 0.097$), suggesting equivalent visual and kinesthetic vividness.

---insert Table 2 about here---

Structural validity

Kaiser-Meyer-Olkin criteria for adequate measure of sampling adequacy of > 0.855, with an acceptable minimum of 0.5, and Bartlett's sphericity ($X^2 \geq 479.414$, df ≥ 91, p < 0.001) tests indicated it was pertinent to use Factor Analysis for the MIQ-RS, KVIQ-20 and KVIQ-10 questionnaires. However, it was not possible for the KVIQ-34 due to excessively high collinearity between items, and thus its psychometric assessment was performed only as a total score and not as two-dimensional tool. Two factors explained the 70.63%, 67.43% and 65.08% of total variance for the MIQ-RS, KVIQ-20 and KVIQ-10 questionnaires, respectively, therefore showing evidence of their two-dimensional underlying structure. These factors were related between each other with r = 0.6-0.63 what indicated that a non-orthogonal rotation was needed. Visual items were grouped together with Factor 1, which explained 53.5-58.73% of variance, and kinesthetic items did the same with Factor 2, which explained the 11.59-11.9% of variance. Mean individual item scores and their factor loadings (each KVIQ version independently) are shown in Table 3.

---insert Table 3 about here---

Reliability

Internal consistency, test-retest reliability, SEM and SDC results are reported in Table 4. Wilcoxon signed-rank tests for all questionnaires and their subscales (when pertinent) confirmed there were not statistically significant differences between test and retest scores (p > 0.05), suggesting that a learning effect was absent. Revelle's Omega total values ≥ 0.9 for the overall scores and subscales (when pertinent) provided evidence of excellent internal consistency. ICC = 0.59-0.75 showed moderate test-retest reliability. The visual subscales of all questionnaires exhibited higher ICCs than their kinesthetic

counterparts. Bland-Altman plots for the KVIQ-20 and MIQ-RS questionnaires are shown in Figure 1.

---insert Table 4 about here---

---insert Figure 1 about here---

Criterion validity

The overall and subscale scores of the KVIQ-20 were statistically significant ($p < 0.01$), positively and strongly ($r > 0.7$) correlated to their MIQ-RS counterparts (Figure 2). The same was found for the KVIQ-10 and the MIQ-RS correlations, thus showing evidence of strong criterion validity between these measures. The KVIQ-10 was strongly and significantly correlated to the KVIQ-20 for the overall, visual and kinesthetic scores ($r=0.97$, $r=0.98$ and $r=0.96$, respectively; $p < 0.01$). The overall score of the KVIQ-34 was correlated to the MIQ-RS, KVIQ-20 and KVIQ-20 overall scores ($r=0.82$, $r=0.99$ and $r=0.98$; $p < 0.01$).

---insert Figure 2 about here---

Discussion

In this study we provide evidence of the validity (structural and criterion) and reliability (internal consistency, test-retest, measurement error) of the Spanish MIQ-RS and KVIQ questionnaires in people with PD. To the best of our knowledge this is the first study that assesses the suitability of the MIQ-RS in this movement disorder. Our results showed evidence of a good psychometric behaviour of the MIQ-RS and the long and short versions of the KVIQ, in comparison to its extended version whose factor structure could not be confirmed.

Our observations are in accordance with previous studies showing evidence of the two-dimensional structure of these instruments in people with stroke and Multiple Sclerosis.[15,22] Here, we further found that the Spanish short and long KVIQ versions as well as the MIQ-RS showed equivalent structural validity and excellent internal consistency in people with PD. However, in our study, it was only possible to extract the two theoretical factors in the long and short KVIQ versions, since very high collinearity between items was found in the extended version. This reflected that some dominant or non-dominant upper and lower limb items were nearly totally predicted by the others, contained in the KVIQ-20, and therefore in this version the score should not be interpreted as two separate subscales but as an overall score.[47] The problem of high collinearity is a well-known limitation of long psychometric questionnaires measuring closely related constructs,[48] where some redundancy is expected between items.[49] This was illustrated by Revelle's Omega total ≥ 0.95 for the total and visual KVIQ-20/MIQ-RS and total KVIQ-34 scores in our study. In this context, perhaps the use of the KVIQ-10 is recommendable as there is evidence of its appropriate factor structure and excellent internal consistency as well as good test-retest reliability and similar measurement error.

Test-retest reliability assessment yielded lower ICCs than previous studies which analysed the English KVIQ-34 or the German KVIQ-20/-10 in people with PD.[17,18] The English version accomplished almost excellent reliability for the total, visual and kinesthetic scores (ICC > 0.8), whereas the German version showed lower data for the KVIQ-20 (ICC = 0.86, 0.68 and 0.82, respectively) or the KVIQ-10 subscales (ICC = 0.69 (visual) and 0.84 (kinesthetic)). However, our results showed ICCs ranging from 0.58 to 0.69 which substantially differ from these previous studies. While the non-normal distribution of our data might have influenced the values of the ICC,[42] other

292 factors might explain these inconsistencies as well. For instance, previous work used a
293 one-way random-effects model of the ICC but a two-way mixed-effects one should be
294 used when assessing test-retest reliability.[33] Additionally, the type and definition of the
295 ICC was not specified, which is relevant in the process of selecting and reporting.
296 Moreover, sample sizes were significantly smaller (n=11 and n=8 vs n=35) and
297 participants had different clinical characteristics (e.g., Hoehn and Yahr stage 1-2.5[17] or
298 mean disease duration of 5.2 (3.4) years[18] vs stages 1-4 or disease duration of 8.8 (5.9)
299 years in our study). These elements should be considered when interpreting ICCs of the
300 different translations and versions of the KVIQ in people with PD, which may have
301 been overestimated in previous literature. Conversely, the MIQ-RS showed test-retest
302 reliability similar to previous studies,[23] ICCs ranging from 0.67 to 0.75 which is likely
303 to reflect the consistency of MI questionnaires over time more appropriately.

304 Consistent positive strong correlations between questionnaires suggested high criterion
305 validity, which is in accordance with precedent works.[17,23] Nonetheless, both
306 instruments should not be used interchangeably in clinical practice, as there are
307 substantial differences between them. On the one hand, the KVIQ evaluates simpler and
308 joint-specific movements of the neck, trunk and upper/lower limbs (e.g., neck flexion,
309 hand finger-tapping, hip abduction or foot stepping), though their functional
310 significance is reduced. However, the MIQ-RS assesses functionally meaningful
311 movements but with an almost exclusive focus on upper limb tasks (e.g., grasping a
312 glass and lifting it off a table or opening a door). On the other hand, in our experience
313 the administration of the KVIQ long and extended versions takes up to thirty minutes
314 but the MIQ-RS and the KVIQ short version need over ten minutes, which is of
315 importance when evaluating their applicability to clinical practice. These differences

should be considered when administering them to people with PD, as well as the psychometric properties obtained in this study.

The results obtained in this work should be interpreted considering the clinical characteristics of the participants enrolled. For instance, PIGD motor phenotype of PD was present in most participants (n=25), and this fact might have impacted on the psychometric analyses considering there is evidence suggesting that tremor may modify MI by modulating central somatosensory processing.[50] Nevertheless, the effect of PD motor phenotype (PIGD or TD) on imagery ability remains largely unexplored. Additionally, despite participants were in a broad range of Hoehn and Yahr stages (1-4), mild to moderate stages (1.5-3) were predominant (n=31). Specifically, most participants were in stages 2.5 or 3 (n=21), which is of importance considering that MI ability seems to correlate with disease severity (measured with the MDS-UPDRS) in "on" medication states.[51] However, there is parallel evidence showing non-significant correlations between measures of imagery ability and Hoehn and Yahr scale.[24,52] Future lines of work with larger sample sizes may be opened in light of these inconsistencies.

Study limitations

Due to the specific clinical characteristics of the participants regarding their motor phenotype or Hoehn and Yahr stage, the generalisability of our results is limited. Some methodological limitations should also be considered. First, we intentionally used a counterbalanced order in the second assessment day to avoid learning effects, but this fact could have affected test-retest reliability producing an order effect. In addition, we used ICCs to assess test-retest reliability despite our data showed non-normal distributions across all variables studied. Because ICC assumes data to be normally

distributed, this may have affected the results.[53] Nonetheless, there is currently no international consensus on alternatives to the ICC in such cases.[42]

Second, questionnaires were administered in the "on" medication state, which limits the applicability of our results, despite motor fluctuations appear to not significantly modify MI ability.[51] Finally, even though our sample size was sufficient to evaluate test-retest reliability,[32] it should be considered small by the requirements of the COSMIN guidelines for assessing structural validity using Factor Analysis.[28] However, Exploratory Factor Analysis can be used in studies with $n < 50$ under some conditions that apply to this work, such as high factor loadings, low number of factors or high number of variables.[54]

Clinical implications

Our results have several contributions to clinical practice. MI ability assessment in Spanish-speaking people with PD can now be conducted accurately using two easily administered, simply interpretable and inexpensive measures. This is a major advantage for clinicians as other assessment methods of imagery ability (i.e., neuroimage techniques or electrophysiological measures) require complex procedures, expensive equipment and extensive experience by the examiner, which limits their clinical usability, applicability and interpretability. Furthermore, SEMs and SDCs of the questionnaires are provided. Therefore, clinical practitioners aiming to use Mental Representation techniques (MI or Action Observation) can evaluate MI ability with these tools and determine precisely how this capacity changes with training programmes.

MI modality-specific vividness can be independently assessed with these measures, that are able to distinguish between Visual and Kinesthetic imagery. Importantly, it is not

mandatory to complete the whole questionnaires, as subscales can be administered in isolation, which may be of interest in specific clinical contexts. Last, there is evidence of the validity and reliability of the KVIQ short version (KVIQ-10), which can be now administered to people with PD and it is significantly less time-consuming than its long counterpart.

Conclusions

The results provide evidence of the validity and reliability of the Spanish MIQ-RS and KVIQ long and short versions to assess MI vividness in people with PD. The KVIQ extended version (KVIQ-34) should be used cautiously because there is no evidence of an adequate structural validity in this population.

Acknowledgements

We want to thank all the participants for their time and willingness. We want to thank Dr. Carmen Bravo-Llatas and Rubén Fernández-Matías, PT, for their statistical advice.

Supplemental Digital Content

Supplemental Digital Content 1. Video Abstract.

References

1. Dana A, Gozalzadeh E. Internal and External Imagery Effects on Tennis Skills Among Novices. *Percept Mot Skills*. 2017;124(5):1022-1043. doi:10.1177/0031512517719611

2. Slimani M, Bragazzi NL, Tod D, et al. Do cognitive training strategies improve motor and positive psychological skills development in soccer players? Insights from a systematic review. *J Sports Sci*. 2016;34(24):2338-2349. doi:10.1080/02640414.2016.1254809

3. Keller PE. Mental imagery in music performance: Underlying mechanisms and potential benefits. *Ann N Y Acad Sci*. 2012;1252(1):206-213. doi:10.1111/j.1749-6632.2011.06439.x

4. Ruffino C, Papaxanthis C, Lebon F. Neural plasticity during motor learning with motor imagery practice: Review and perspectives. *Neuroscience*. 2017;341:61-78. doi:10.1016/j.neuroscience.2016.11.023

5. Abraham A, Hart A, Dickstein R, Hackney ME. "Will you draw me a pelvis?" Dynamic neuro-cognitive imagery improves pelvic schema and graphic-metric representation in people with Parkinson's Disease: A randomized controlled trial. *Complement Ther Med*. 2019;43:28-35. doi:10.1016/j.ctim.2018.11.020

6. Tamir R, Dickstein R, Huberman M. Integration of motor imagery and physical practice in group treatment applied to subjects with Parkinson's disease. *Neurorehabil Neural Repair*. 2007;21(1):68-75. doi:10.1177/1545968306292608

7. El-Wishy AA, Fayez ES. Effect of locomotor imagery training added to physical therapy program on gait performance in Parkinson patients: A randomized controlled study. *Egypt J Neurol Psychiatry Neurosurg*. 2013;50(1):31-37.

8. Braun S, Beurskens A, Kleynen M, Schols J, Wade D. Rehabilitation with mental practice has similar effects on mobility as rehabilitation with relaxation in people with Parkinson's disease: A multicentre randomised trial. *J Physiother*. 2011;57(1):27-34. doi:10.1016/S1836-9553(11)70004-2

9. De Melo Santiago LM, De Oliveira DA, De Macêdo Ferreira LGL, et al. Immediate effects of adding mental practice to physical practice on the gait of individuals with Parkinson's disease: Randomized clinical trial. *NeuroRehabilitation*. 2015;37(2):263-271. doi:10.3233/NRE-151259

10. Ruffino C, Papaxanthis C, Lebon F. The influence of imagery capacity in motor performance improvement. *Exp Brain Res*. 2017;235(10):3049-3057. doi:10.1007/s00221-017-5039-8

11. Collet C, Guillot A, Lebon F, MacIntyre T, Moran A. Measuring Motor Imagery Using Psychometric, Behavioral, and Psychophysiological Tools. *Exerc Sport Sci Rev*. 2011;39(2):85-92. doi:10.1097/JES.0b013e31820ac5e0

12. Williams SE, Guillot A, Di Rienzo F, Cumming J. Comparing self-report and

mental chronometry measures of motor imagery ability. *Eur J Sport Sci*. 2015;15(8):703-711. doi:10.1080/17461391.2015.1051133

13. Hanakawa T. Organizing motor imageries. *Neurosci Res*. 2016;104:56-63. doi:10.1016/j.neures.2015.11.003

14. Melogno-Klinkas M, Núñez-Nagy S, Ubillos-Landa S. Revisión sistemática sobre instrumentos de valoración de la imaginación motora para población hispanohablante: Su uso en rehabilitación. *Rev Neurol*. 2017;65(9):385-395.

15. Malouin F, Richards CL, Jackson PL, Lafleur MF, Durand A, Doyon J. The Kinesthetic and Visual Imagery Questionnaire (KVIQ) for Assessing Motor Imagery in Persons with Physical Disabilities: A Reliability and Construct Validity Study. *J Neurol Phys Ther*. 2007;31:20-29. doi:10.1097/01.NPT.0000260567.24122.64

16. Gregg M, Hall C, Butler A. The MIQ-RS: A suitable Option for examining movement imagery ability. *Evidence-based Complement Altern Med*. 2010;7(2):249-257. doi:10.1093/ecam/nem170

17. Randhawa B, Harris S, Boyd LA. The kinesthetic and visual imagery questionnaire is a reliable tool for individuals with Parkinson Disease. *J Neurol Phys Ther*. 2010;34(3):161-167. doi:10.1097/NPT.0b013e3181e1aa71

18. Schuster C, Lussi A, Wirth B, Ettlin T. Two assessments to evaluate imagery ability: Translation, test-retest reliability and concurrent validity of the German KVIQ and Imaprax. *BMC Med Res Methodol*. 2012;12:127-140. doi:10.1186/1471-2288-12-127

19. Demanboro A, Sterr A, dos Anjos SM, Conforto AB. A Brazilian-Portuguese version of the kinesthetic and visual motor imagery questionnaire. *Arq Neuropsiquiatr*. 2018;76(1):26-31. doi:10.1590/0004-282x20170181

20. Nakano H, Kodama T, Ukai K, Kawahara S, Horikawa S, Murata S. Reliability and validity of the Japanese version of the kinesthetic and visual imagery questionnaire (KVIQ). *Brain Sci*. 2018;8(5). doi:10.3390/brainsci8050079

21. Loison B, Moussaddaq AS, Cormier J, et al. Translation and validation of the French Movement Imagery Questionnaire - Revised Second version (MIQ-RS). *Ann Phys Rehabil Med*. 2013;56(3):157-173. doi:10.1016/j.rehab.2013.01.001

22. Moghadas Tabrizi Y, Zangiabadi N, Mazhari S, Zolala F. The reliability and validity study of the Kinesthetic and Visual Imagery Questionnaire in individuals with Multiple Sclerosis. *Brazilian J Phys Ther*. 2013;17(6):588-592. doi:10.1590/S1413-35552012005000124

23. Butler AJ, Cazeaux J, Fidler A, et al. The movement imagery questionnaire-revised, second edition (MIQ-RS) is a reliable and valid tool for evaluating motor imagery in stroke populations. *Evidence-based Complement Altern Med*. 2012;2012(497289). doi:10.1155/2012/497289

24. Heremans E, Feys P, Nieuwboer A, et al. Motor imagery ability in patients with

early- and mid-stage Parkinson disease. *Neurorehabil Neural Repair*. 2011;25(2):168-177. doi:10.1177/1545968310370750

25. Heremans E, D'Hooge AM, De Bondt S, Helsen W, Feys P. The relation between cognitive and motor dysfunction and motor imagery ability in patients with multiple sclerosis. *Mult Scler J*. 2012;18(9):1303-1309. doi:10.1177/1352458512437812

26. Abraham A, Hart A, Andrade I, Hackney ME. Dynamic neuro-cognitive imagery improves mental imagery ability, disease severity, and motor and cognitive functions in people with Parkinson's disease. *Neural Plast*. 2018;2018(6168507). doi:10.1155/2018/6168507

27. Maillet A, Thobois S, Fraix V, et al. Neural substrates of levodopa-responsive gait disorders and freezing in advanced Parkinson's disease: A kinesthetic imagery approach. *Hum Brain Mapp*. 2015;36(3):959-980. doi:10.1002/HBM.22679

28. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737-745. doi:10.1016/j.jclinepi.2010.02.006

29. Williams SE, Cumming J, Ntoumanis N, Nordin-Bates SM, Ramsey R, Hall C. Further validation and development of the movement imagery questionnaire. *J Sport Exerc Psychol*. 2012;34(5):621-646. doi:10.1123/jsep.34.5.621

30. Cantalejo-Fernández M, Díaz-Arribas MJ, Fernández-de-las-Peñas C, Plaza-Manzano G, Ríos-León M, Martín-Casas P. Translation and validation of the Spanish Movement Imagery Questionnaire Revised Second version (MIQ-RS). *PM&R*. 2021;10.1002/pm. doi:10.1002/pmrj.12546

31. Melogno-Klinkas M, Ubillos-Landa S, Nuñez-Nagy S. Kinesthetic and visual motor imagery in Spanish healthy and stroke people. *PsyArxiv Prepr*. 2022;March 30. doi:10.31234/OSF.IO/VJ2RE

32. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat Med*. 2012;31(29):3972-3981. doi:10.1002/sim.5466

33. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

34. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: A clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry*. 1992;55(3):181-184. doi:10.1136/jnnp.55.3.181

35. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-198. doi:10.1016/0022-3956(75)90026-6

36. McKay MJ, Baldwin JN, Ferreira P, et al. Normative reference values for strength and flexibility of 1,000 children and adults. *Neurology*. 2017;88(1):36. doi:10.1212/WNL.0000000000003466

37. Norkin CC, White DJ, Torres J, et al. *Measurement of Joint Motion: A Guide to Goniometry*. 5th ed. F.A. Davis Company; 2016.

38. Stebbins GT, Goetz CG, Burn DJ, Jankovic J, Khoo TK, Tilley BC. How to identify tremor dominant and postural instability/gait difficulty groups with the movement disorder society unified Parkinson's disease rating scale: Comparison with the unified Parkinson's disease rating scale. *Mov Disord*. 2013;28(5):668-670. doi:10.1002/mds.25383

39. Mena MÁ, Casarejos MJ, Solano RM, García De Yébenes J. Half a century of L-DOPA. *Curr Top Med Chem*. 2009;9(10):880-893. doi:10.2174/156802609789378263

40. Cho E. *Making Reliability Reliable: A Systematic Approach to Reliability Coefficients*. Vol 19.; 2016. doi:10.1177/1094428116656239

41. Mcneish D. Thanks Coefficient Alpha, We'll Take It From Here. *Psychol Methods*. 2018;23(3):412-433. doi:10.1037/met0000144

42. Mokkink LB, Boers M, van der Vleuten CPM, et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med Res Methodol*. 2020;20(1):1-13. doi:10.1186/S12874-020-01179-5/TABLES/7

43. Gerke O. Reporting standards for a bland-altman agreement analysis: A review of methodological reviews. *Diagnostics*. 2020;10(5). doi:10.3390/diagnostics10050334

44. Furlan L, Sterr A. The applicability of standard error of measurement and minimal detectable change to motor learning research - A behavioral study. *Front Hum Neurosci*. 2018;12. doi:10.3389/fnhum.2018.00095

45. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157. doi:10.1007/S11136-018-1798-3

46. Schober P, Boer C, Schwarte LA. Correlation Coefficients. *Anesth Analg*. 2018;126(5):1763-1768. doi:10.1213/ANE.0000000000002864

47. Schreiber JB. Issues and recommendations for exploratory factor analysis and principal component analysis. *Res Soc Adm Pharm*. Published online 2020. doi:10.1016/j.sapharm.2020.07.027

48. Dormann CF, Elith J, Bacher S, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography (Cop)*. 2013;36(1):27-46. doi:10.1111/j.1600-0587.2012.07348.x

49. Streiner DL. Starting at the beginning: An introduction to coefficient alpha and

internal consistency. *J Pers Assess*. 2003;80(1):99-103.

50. Helmich RC, Bloem BR, Toni I. Motor imagery evokes increased somatosensory activity in parkinson's disease patients with tremor. *Hum Brain Mapp*. 2012;33(8):1763-1779. doi:10.1002/hbm.21318

51. Peterson DS, Pickett KA, Earhart GM. Effects of levodopa on vividness of motor imagery in Parkinson disease. *J Parkinsons Dis*. 2012;2(2):127-133. doi:10.3233/JPD-2012-12077

52. Pickett, Kristen A. Peterson, Daniel S. Earhart GM. Motor imagery of gait tasks in individuals with Parkinson disease. *J Parkinsons Dis*. 2012;2(1):19-22. doi:10.3233/JPD-2012-11045

53. Mehta S, Bastero-Caballero RF, Sun Y, et al. Performance of intraclass correlation coefficient (ICC) as a reliability index under various distributions in scale reliability studies. *Stat Med*. 2018;37(18):2734. doi:10.1002/SIM.7679

54. F de Winter JC, Dodou D, Wieringa PA. Exploratory Factor Analysis With Small Sample Sizes. *Multivariate Behav Res*. 2009;44(2):147-181. doi:10.1080/00273170902794206

Tables and Figure captions

Table 1. Functional status of the participants.

Variable	Category	Mean score	SD
MDS-UPDRS	Total	59.71	18.63
	Part I	13.14	6.53
	Part II	16.17	6.85
	Part III	26.43	8.19
	Part IV	3.97	3.34
Motor Phenotype	TD	n = 9 (25.71%)	
	PIGD	n = 25 (71.43%)	
	Indeterminate	n = 1 (2.9%)	
Hoehn and Yahr Stage	1	n = 1 (2.9%)	
	1.5	n = 5 (14.3%)	
	2	n = 5 (14.3%)	
	2.5	n = 8 (22.9%)	
	3	n = 13 (37.1%)	
	4	n = 3 (8.6%)	
Schwab & England scale	40%	n = 1 (2.9%)	
	60%	n = 3 (8.6%)	
	70%	n = 4 (11.4%)	
	80%	n = 11 (31.4%)	
	90%	n = 15 (42.9%)	
	100%	n = 1 (2.9%)	

Abbreviations: MDS-UPDRS, Movement Disorders Society-Unified Parkinson's Disease Rating Scale; PD, Parkinson's Disease; PIGD, Postural Instability/Gait Difficulty; TD, Tremor Dominant.

562 Table 2. Total, visual and kinesthetic scores of the Spanish KVIQ (long, short and

563 extended versions) and MIQ-RS questionnaires.

	All participants (n=35)		TD Phenotype (n=9)		PIGD Phenotype (n=25)	
Questionnaire	Mean	SD	Mean	SD	Mean	SD
KVIQ-20 Total	72.86	20.52	75.79	20.92	71.04	20.71
KVIQ-20 Visual	36.23	12.27	37.01	15.72	35.56	11.24
KVIQ-20 Kinesthetic	36.63	10.43	38.78	10.2	35.48	10.62
KVIQ-10 Total	37.06	10.15	38.67	9.41	36.2	10.62
KVIQ-10 Visual	18.43	6.05	18.78	7.69	18.16	5.61
KVIQ-10 Kinesthetic	18.63	5.31	19.89	4.46	18.04	5.65
KVIQ-34 Total	125.4	34.53	129.34	34.46	122.8	35.27
KVIQ-34 Visual	62.6	20.89	63.33	26.93	61.2	19.13
KVIQ-34 Kinesthetic	62.8	17.44	66.01	16.82	61.08	17.92
MIQ-RS Total	72.91	19.59	74.01	12.85	72.13	21.98
MIQ-RS Visual	37.00	11.33	36.11	13.01	37.12	11.14
MIQ-RS Kinesthetic	35.91	10.35	37.9	6.95	35.01	11.49

564

565 Note: One participant was categorized as “indeterminate” according to Motor

566 Phenotype classification and their data is not shown in this table. Abbreviations: KVIQ-

567 (10/20/34), Kinesthetic and Visual Imagery Questionnaire-(short/long/extended); MIQ-

568 RS, Movement Imagery Questionnaire-Revised Second Edition; PIGD, Postural

569 Instability/Gait Difficulty; TD, Tremor Dominant.

570 Table 3. Mean score of the individual items for the KVIQ-20, KVIQ-10 and MIQ-RS questionnaires and their factor loadings.

KVIQ-20					KVIQ-10					MIQ-RS				
Item	Mean	SD	Factor		Item	Mean	SD	Factor		Item	Mean	SD	Factor	
			1	2				1	2				1	2
Item 10Vd	3.8	1.45	0.941	-	Item 8Vd	3.74	1.42	0.91	-	Item 13 (V)	4.91	1.77	0.960	-
Item 7Vnd	3.83	1.29	0.936	-	Item 9Vnd	3.54	1.44	0.864	-	Item 10 (V)	5.23	1.76	0.902	-
Item 8Vd	3.74	1.42	0.876	-	Item 5Vd	3.91	1.31	0.741	-	Item 4 (V)	5.11	1.74	0.897	-
Item 9Vnd	3.54	1.44	0.865	-	Item 6V	3.66	1.43	0.729	-	Item 8 (V)	5.00	1.79	0.886	-
Item 2V	3.4	1.44	0.847	-	Item 3Vnd	3.57	1.5	0.722	-	Item 2 (V)	5.23	2.00	0.818	-
Item 3Vnd	3.57	1.5	0.819	-	Item 5Kd	3.69	1.28	-	0.974	Item 5 (V)	5.14	1.88	0.774	-
Item 1V	3.26	1.48	0.809	-	Item 8Kd	3.97	1.2	-	0.829	Item 14 (V)	4.89	1.90	0.753	-
Item 5Vd	3.91	1.31	0.754	-	Item 9Knd	3.71	1.2	-	0.793	Item 9 (K)	5.34	1.76	-	0.897
Item 4Vd	3.51	1.54	0.734	-	Item 6K	3.63	1.29	-	0.714	Item 6 (K)	5.46	1.68	-	0.864
Item 6V	3.66	1.43	0.694	-	Item 3Knd	3.63	1.46	-	0.505	Item 7 (K)	5.43	1.78	-	0.858
Item 10Kd	3.86	1.22	-	0.985	Mean	3.71	1.35	-	-	Item 11 (K)	5.11	1.85	-	0.828
Item 5Kd	3.69	1.28	-	0.922						Item 12 (K)	5.29	1.65	-	0.811
Item 8Kd	3.97	1.2	-	0.88						Item 3 (K)	5.37	1.80	-	0.587
Item 9Knd	3.71	1.2	-	0.808						Item 1 (K)	5.40	1.80	-	0.569
Item 7Knd	3.86	1.22	-	0.772						Mean	5.20	1.80	-	-
Item 6K	3.63	1.29	-	0.76										
Item 4Kd	3.51	1.31	-	0.75										
Item 1K	3.37	1.33	-	0.664										

Item 2K	3.4	1.46	-	0.555	
Item 3Knd	3.63	1.46	-	0.537	
Mean	3.64	1.36	-	-	

571 Factor loadings < 0.2 have been removed. Abbreviations: V, Visual; K, Kinesthetic; d, Dominant side; nd, non-Dominant side.

572 Table 4. Reliability results of the total, visual and kinesthetic scores for the KVIQ

573 (short, long and extended versions) and MIQ-RS questionnaires.

	Revelle's Omega total coefficient	Day 1		Day 2		ICC (95% CI)	SEM	SDC
		Mean	SD	Mean	SD			
KVIQ-20 Total	0.97	72.86	20.52	78.94	16.33	0.663 (0.410-0.818)	11.92	33.02
KVIQ-20 Visual subscale	0.96	36.23	12.27	40.02	9.33	0.656 (0.395-0.815)	7.19	19.94
KVIQ-20 Kinesthetic subscale	0.94	36.63	10.43	38.91	8.73	0.585 (0.323-0.765)	6.72	18.63
KVIQ-10 Total	0.94	37.06	10.15	39.94	8.17	0.673 (0.428-0.823	5.56	15.41
KVIQ-10 Visual subscale	0.91	18.43	6.05	20.17	4.64	0.674 (0.425-0.824)	3.3	9.14
KVIQ-10 Kinesthetic subscale	0.9	18.63	5.31	19.77	4.49	0.604 (0.349-0.777)	3.3	9.14
KVIQ-34 Total	0.97	125.4	34.53	134.06	27.31	0.692 (0.463-0.833)	19.16	53.12
MIQ-RS Total	0.97	72.91	19.59	71.77	17.63	0.681 (0.452-0.825)	11.15	30.9
MIQ-RS Visual Subscale	0.96	37.00	11.33	36.40	10.34	0.752 (0.562-0.867)	5.71	15.83
MIQ-RS Kinesthetic Subscale	0.92	35.91	10.34	35.37	9.37	0.671 (0.437-0.819)	5.99	16.59

574

575 Abbreviations: ICC, Intraclass Correlation Coefficient; SDC, Smallest Detectable

576 Change; SEM, Standard Error of Measurement.

Figure 1. Bland-Altman plots for the total, visual and kinesthetic scores of the Spanish KVIQ-20 and MIQ-RS questionnaires. Panels A-C show plots for the MIQ-RS and D-F for the KVIQ-20. Each panel show the difference between test and retest plotted against the mean of the two days. The mean difference (continuous line) and upper and lower limits for a 95% confidence interval (discontinuous lines) are shown in blue, red and green for the total, visual and kinesthetic scores of both questionnaires, respectively.

Figure 2. Spearman's rank correlation coefficients between the total and subscale scores (if pertinent) of the Spanish KVIQ and MIQ-RS questionnaires.