# Chapter 4

# Evidence for Reliability of the DMQ

*George A. Morgan, Su-Ying Huang, Stephen Amukune, Jessica M. Gerton,
Ágnes Nyitrai and Krisztián Józsa*

## Introduction

This chapter provides data about evidence for the measurement reliability of DMQ 18, which builds on evidence from DMQ 17. There are several methods to assess measurement reliability: internal consistency, test-retest, interrater, and parallel forms. We end the chapter with an extended conclusion, which provides summary statements about evidence for the reliability of the DMQ. The next section will describe and define the several methods that provide evidence for the reliability of a questionnaire such as the DMQ.

## Types of Evidence to Support Reliability

Reliability refers to the consistency of a measure within a scale, over time, or among raters. Reliability is essential for a measure to be valid because if a measure is inconsistent, it cannot be a good or valid measure of the construct to be assessed. Several types of evidence have been used in the litera-

ture provide support for the reliability of a measure. There are three common types of evidence for evaluating the reliability of a measure: internal consistency, test-retest reliability, and interrater reliability.

### Internal Consistency Measures

The most common measure of internal consistency is coefficient alpha, popularized by Cronbach (1951) and referred to in this book as **Cronbach alpha**, which is based on the intercorrelations of the several ratings that are used to develop a summary measure or scale. In the DMQ, there are 6 motivation scales: Cognitive/Object Persistence, Gross Motor Persistence, Social Persistence with Adults, Social Persistence with Children, Mastery Pleasure, and Negative Reactions to Challenge (which was intended to have two subscales). Each scale has several items rated on Likert scales from 1-5. Cronbach alphas are almost always used to test the internal consistency of a set of Likert scale items that form a composite scale. If the items in a summary or composite scale are highly correlated, the Cronbach alpha will be positive and high, approaching 1.0. The Cronbach alpha coefficient depends heavily on the number of items in the scale, so that with two or only a few items, a high alpha may be difficult to obtain, unless the items are highly intercorrelated. With 10 or more items, alphas are almost always above .80, unless there are very low or negative correlations among some pairs of the items. If there are negative correlations among the items, one should be careful to make sure that the items are all coded in the same direction so that a high score on every item would mean the same thing (e.g., high Gross Motor Persistence or high Mastery Pleasure). If there are negatively worded items in the scale, they would need to be reverse coded so that a high rating would indicate the same thing on each item.

Cronbach alphas also can be used with true/false or right/wrong questions (dichotomous scores), but that is relatively uncommon. There are also other statistics to assess internal consistency reliability, such as split-half methods using the Spearman-Brown formula, that are more useful if the items are dichotomous.

### Test-Retest Reliability

Test-retest reliability assesses consistency in the ratings of the same group of persons over a short period of time, from a week to a month or so. Both internal consistency and test-retest reliability can use a correlation coefficient or an intraclass correlation coefficient (ICC) to assess whether there is consistency in the ratings. When the period of time is a week or two, the correlation coefficient or ICC is often high, $r \geq .80$. With the ICC, one also gets a test of statistical significance, but this test only indicates whether the ICC coefficient is greater than 0, so usually not very important.

### *Interrater Reliability*

This type of reliability measure is used when two or more different raters rate the same subject, such as a child rated with the DMQ by two teachers, to assess the extent to which the raters agree. Again, this could be done with a correlation coefficient or the ICC. The latter is especially useful when there are more than two raters. Again, the coefficient should be positive and high, ≥ .70.

With the DMQ, it is difficult to find situations where the interrater reliability is appropriate. If, as in a couple of the preschool studies, there are two teachers who see the kids at the same or somewhat overlapping times of day, there may be an appropriate measure of interrater reliability. However, often we have self-ratings by the child and a teacher rating, or a rating by the child and a parent rating of the same child. These ratings are in somewhat different contexts because the child is not in school the whole day and is not home with the parent the whole day. Thus, we would not expect the teacher, parent, and child ratings to be highly correlated. We have considered such ratings to be evidence of the construct validity of the measure, rather than a measure of test-retest reliability. (See **Chapter 6**)

### *Parallel Forms Reliability*

Another type of test for reliability is called parallel forms reliability. With standardized tests, there is often more than one version or form of the instrument that presumably measures the same concept. There is only one version of DMQ 17 or 18 for each language so, we cannot test for parallel forms reliability with the same DMQ version in the same language. However, somewhat similar to parallel forms is the situation where persons rated both DMQ 17 and 18, or rated the English and a local language version of the form.

In this chapter, Cronbach alphas, ICCs, and correlation coefficients of .70 and above were judged to be acceptable; equal to or greater than .80 is good. Alphas .60-.69 were said to be minimally acceptable. Those below .60 are low and usually considered unacceptable. Negative coefficients indicate some type of error.

### *Empirical Evidence for Reliability in This Chapter*

Evidence supporting the reliability of DMQ 18 is accumulating. Evidence about the reliability is also available from DMQ 17, which has the same scales and similar items. DMQ 17 evidence will be summarized first, as background for DMQ 18. In general, the current DMQ 18 data show similar reliabilities to the earlier version. We expect that other DMQ 18 data being collected in the future will provide further support for the reliability of DMQ 18. Following the summary of internal consistency reliability for DMQ 17,

we divided the discussion of internal consistency for DMQ 18 into preschool (with a couple of infant samples) in Table 4.2 and school age in Table 4.3.

# Internal Consistency Reliability

### *Summary of Internal Consistency for DMQ 17*

Although there were a number of individual studies that provided evidence for the internal consistency of DMQ 17 and earlier versions, summary chapters by Morgan et al. (2013) and by Józsa and Molnár (2013) provided alphas for pooled DMQ 17 samples.

Table 4.1 presents alpha reliability evidence for the 6 mastery motivation scales from large, mixed-age datasets separately pooled from several English language studies and from several Chinese language studies, after excluding the negatively worded item from each scale. The table indicates that the four DMQ 17 persistence or **instrumental scales** and the **Mastery Pleasure Scale** had acceptable to good internal consistency (alphas > .74) for both English and Chinese parent versions and also for the English version rated by teachers. Alphas for the child self-ratings were somewhat lower (.67 - .85) on these five scales. Alphas for the **Negative Reactions to Failure** scale for DMQ 17 also were lower than for the persistence scales. Namely, alphas for the Negative Reactions to Failure scale ranged from .60 - .82, median .65 (Morgan et al., 2013). These lower reliabilities for the Negative Reactions to Failure scale were one reason that DMQ 17 was revised to create DMQ 18. A second reason was that some of the social persistence items seemed to be less appropriate for school age children than for younger children, especially when rated by the children themselves.

Some of the English-speaking children in the Morgan et al. (2013) data were 5-7 years old, probably too young to understand fully these self-ratings of their motivation, even when the items were read to them and/or the tester used visual aids. These young school-aged children had the lowest alphas (.61 - .85, median .68). Gilmore and Boulton-Lewis (2009) in Australia also found lower alphas from self-ratings by their young school-age children. Seventeen out of 20 of their 8-year-olds had a variety of learning disabilities, which also may have led to difficulties in making such self-ratings.

Józsa and Molnár (2013) and Józsa et al. (2014) reported on several studies with large Hungarian samples of school-age children and found acceptable (.67-.84, median .76) Cronbach alphas for the four persistence scales and Mastery Pleasure for self-ratings by children. Alphas for teachers' and parents' ratings of the child were also acceptable and somewhat higher. Reliabilities of those Hungarian teacher ratings were somewhat higher than the alphas for parents. Józsa did not provide information about the Negative Reactions to Failure scale.

*Table 4.1. DMQ 17 Cronbach Alphas for Composite English and Chinese Samples*

| DMQ scales | Parent ratings | | | | Teacher | Child-self | |
|---|---|---|---|---|---|---|---|
| *N=* | TE 894 | TC 769 | AE 176 | AC 101 | TE 363 | TE 199 | TC 611 |
| **Instrumental/persistence scales** | | | | | | | |
| **Object Oriented Persistence** | .85 | .76 | .86 | .85 | .91 | .78 | .75 |
| **Gross Motor Persistence** | .89 | .83 | .90 | .82 | .91 | .85 | .85 |
| **Social Persistence with Adults** | .78 | .74 | .79 | .79 | .85 | .68 | 82 |
| **Social Persistence with Children** | .83 | .80 | .89 | .89 | .88 | .67 | .76 |
| **Expressive/affective scales** | | | | | | | |
| **Mastery Pleasure** | .86 | .75 | .91 | .87 | .88 | .80 | .79 |
| **Negative Reactions to Failure** | .73 | .64 | .71 | .65 | .82 | .63 | .60 |

*Note. AC = Atypical Chinese-speaking; AE = Atypical English-speaking; TC = Typical Chinese-speaking; TE = Typical English-speaking; adapted from Morgan, et al. (2013).*

No significant *age* differences in the alpha reliabilities were found for either the teacher or the parent samples. However, reliability for student self-ratings was somewhat higher for older-age groups than younger-age groups. Development of reading comprehension undoubtedly influences the computed reliability of the questionnaire, and it could be the reason for the increase in self-rated reliability coefficients with age.

The summaries from Morgan et al. (2013) and from Józsa and Molnár (2013) provide evidence for the internal consistency of DMQ 17. Alphas for the four instrumental/ persistence scales combined (total persistence) were almost always greater than .80, even for child self-ratings of young children with disabilities. Alphas for teacher ratings were the highest and child self-ratings the lowest, especially for children under age 9. These DMQ 17 alphas across three languages and nationalities encouraged international use. Accordingly, DMQ 18 has been translated into several other languages.

### *Internal Consistency Reliability of the DMQ 18 Scales for Infants and Preschool Children*

The studies shown in Table 4.2 provide Cronbach alpha reliabilities for 18 samples of young children using 9 different languages. The table shows samples that include infants as young as 6 months and preschool children from a variety of countries. (Note that in Kenya and some other countries, children are sometimes allowed to stay in preschool well past the age of 6 years.)

*Table 4.2. Cronbach Alpha Internal Consistency Reliability of the Revised Dimensions of Mastery Motivation Questionnaire (DMQ 18) for Infants and Preschool Children Rated by Parents or Teachers*

| Age | Raters/ language | Child status/ Ns | Instrumental/persistence aspect | | | | Expressive aspect | |
|---|---|---|---|---|---|---|---|---|
| | | | Cognitive/ object | Gross motor | Social w adults | Social w children | Mastery pleasure | Negative reaction |
| 6-10 mo | Par/Eng and Span[a, b] | PT=56 FT=29 | .76 | .69 | .82 | .84 | .74 | .75 |
| 18-20 mo | Par/Eng and Span[c, i] | PT=79 | .84 | .83 | .82 | .84 | .81 | .82 |
| 18-20 mo | Par/Eng and Span[c, i] | FT=37 | .86 | .82 | .75 | .83 | .78 | .79 |
| 1-4½ yr | Par/Chin[a] | TD=45 | .82 | .71 | .85 | .67 | .74 | .79 |
| 1-4½ yr | Par/Chin[a] | DD=40 | .83 | .82 | .81 | .87 | .90 | .81 |
| 1-4 yr | Par/Hun[a] | TD=197 | .84 | .88 | .78 | .84 | .82 | .82 |
| 3-6 yr | Tea/Hun[d] | TD=211 | .93 | .96 | .91 | .90 | .90 | .79 |
| 2-3½ yr | Par/Chin[a] | DD=64 | .84 | .88 | .86 | .75 | .88 | .65 |
| 3-6 yr | Tea/Ban[e] | TD=206 | .89 | .94 | .89 | .88 | .85 | .83 |
| 3-6 yr | Tea/Tur[f] | TD=1592 | .89 | .88 | .88 | .85 | .87 | .80 |
| 3-6 yr | Par/Eng[g] | TD=57 | .80 | .67 | .65 | .84 | .80 | .83 |
| 5-7 yr | Par/Ind[h] | TD=417 | .67 | .71 | .70 | .69 | .90 | - |
| 2-6 ½ yr | Par/Chin[j] | TD=145 | .80 | .80 | .77 | .79 | .76 | .78 |
| 19-42 mo | Par/Chin[k] | SD=56 | .65 | .85 | .75 | .83 | .80 | .76 |
| 31-80 mo | Par/Chin[l] | DD=110 | .84 | .79 | .79 | .87 | .85 | .81 |
| 5-8 yr | Tea/Kis[m] | TD=397 | .83 | .85 | .89 | .89 | .91 | .91 |
| 6-18 mo | Par/Port[n] | TD=20 | .77 | .75 | .82 | .91 | .73 | .87 |
| 2-6 yr | Par/Port[n] | TD=22 | .81 | .72 | .80 | .77 | .69 | .81 |

*Note. Ban = Bangla; Chin = Chinese; DD = developmental delay; Eng = English; FT = full term; Hun = Hungarian; Ind = Indonesian; Kis = Kiswahili; Negative reaction = Negative Reactions to Challenge; Par = Parent; Port = Portuguese; PT = preterm; SD = Speech Delay; Social w adults = Social Persistence with Adults; Social w children = Social Persistence with Children; Span = Spanish; TD = typically developing; Tea = Teacher; Tur = Turkish.*

*[a]Morgan, et al. (2017); [b]Blasco, et al. (2020); [c]Saxton et al. (2020); [d]Józsa & Morgan (2015); [e]Shaoli et al. (2019); [f]Özbey (2020); [g]Wang & Lewis (2019); [h]Rahmawati, et al. (2020); [i]Blasco et al. (2019); [j] Huang & Lo (2019); [k]Chang, et al. (2020); [l]Huang & Chen (2020); [m]Amukune et al. (2020), a few of these preschool children in Kenya were as old at 12 years, but 52% were 5-6 and 86% were 5-8; [n]Brandão et al. (2020)*

Alphas for the persistence scales were all at least minimally acceptable, with only 7 of 72 being minimally acceptable and most being very good, above .80. The minimally accepted alphas were distributed across the four specific persistence scales. Six of the 18 samples included young children at risk or with delay, but there seemed to be little difference in the alphas for children who were at risk or delayed and children developing typically. There also were no clear differences in alphas between the 9 languages. The Turkish, Bangladeshi, Hungarian, and Portuguese samples did not have any minimally acceptable alphas on the persistence scales, and the other language samples had only one or two such alphas. Studies that reported overall (total) persistence found very good alphas, probably because of the increased number of items.

Alphas for the expressive scales, Mastery Pleasure, and overall Negative Reactions to Challenge were acceptable, with only two minimally acceptable alphas (out of 35). All of the other alphas were above .70, and thus acceptable to very good.

Not shown are the alphas for the negative reactions subscales, which varied from unacceptable to good, with the Negative Reactions Sadness/Shame subscale having the lowest, sometimes unacceptable alphas. Thus, revision of the Negative Reactions Sadness/Shame subscale seems necessary before it is used as a separate subscale. Józsa and Barrett's 2018 study with DMQ 17 preschool Hungarian data suggests that some of the negatively worded persistence items, used in the DMQ 17 but not in the DMQ 18, may be useful in such a revision. See the discussion of the Józsa and Barrett study under Evidence for Convergent Validity for the DMQ 17 in **Chapter 5**.

In summary, Cronbach alphas for infants and preschool children indicate that there is acceptable to good internal consistency reliability. This is true for all 6 DMQ 18 scales, in 9 languages and for children with and without developmental disabilities.

### *Internal Consistency for the School-age DMQ 18*

Table 4.3 shows 16 sets of ratings of 8-18 year-old children rated by a parent, teacher, and/or themselves. There were only 13 independent samples for two reasons: the Hungarian 10-11 year-old children were rated by parents, teachers, and the children themselves, the 10-12 year-old Persian-speaking children rated themselves and were rated by a parent. The raters were from five countries, but spoke six languages: Chinese, Hungarian, Persian, Russian, Romanian, or Portuguese. The Russian and Romanian children lived in Moldova. All of the Cognitive/Object Persistence and Gross Motor Persistence alphas were acceptable, but in two of the 32 scales, both in Iran, there was a marginally acceptable Cognitive/Object Persistence sample.

Alphas for the social persistence (mastery motivation) scales were somewhat weaker, with 9 of the 32 scales having marginally acceptable reliability

and 1 scale was unacceptable. Six of these 10 scales were for Persian-speaking raters. The other 20 alphas were acceptable to good.

For Mastery Pleasure, 10 of the 16 alphas had acceptable reliabilities, and the other 6 were marginally acceptable, including all three from the Persian-speaking raters.

For overall Negative Reactions to Challenge, 14 of the 16 had at least minimally acceptable alphas, but two self-rated samples of students from Taiwan had unacceptable alphas. Alphas for the Negative Reactions Sadness/Shame subscale were only minimally acceptable or not acceptable, again supporting the need for revisions.

Thus, it seems that ratings for school-age children had somewhat lower levels of reliability than for infants and preschool children. This seems especially true for self-ratings of these 10-14 year-old children and for most of the scales rated by the Persian-speaking parents and children. There were only two samples of children with disabilities, rated by their parents. Reliabilities for these samples seem similar to those for the other samples of this age group.

Not shown in Table 4.3 is a study of 8-16 children with cerebral palsy by Hines and Bundy (2018), which used only the cognitive persistence scale; they found excellent alphas for their parent ratings.

To summarize the alphas for DMQ 17 and 18, the alphas for the Cognitive/Object Persistence and Gross Motor Persistence scales were acceptable to good for almost all of the samples from the several languages at both preschool and school-age and for children with and without disabilities. All 6 motivation scales had acceptable to good reliability for most preschool DMQ 18 samples; however, reliability was sometimes minimally acceptable and occasionally unacceptable for school-aged samples. Note that the DMQ 18 data are mostly from smaller, single-study samples and from a wide variety of different countries and languages. Samples with exceptions to acceptable alphas usually involved samples of children with disabilities and/or from non- European languages. Further work is needed to understand better cultural and language differences that may underlie these somewhat lower reliabilities.

*Table 4.3. Cronbach Alpha Internal Consistency Reliability of the Revised Dimensions of Mastery Motivation Questionnaire (DMQ 18) for 8-18 Year-old Children*

| Age (years) | Raters/ language | Child status/Ns | Instrumental/persistence aspect | | | | Expressive aspect | |
|---|---|---|---|---|---|---|---|---|
| | | | Cognitive/ object | Gross motor | Social w adults | Social w children | Mastery pleasure | Negative reaction |
| 10-11 | Self/Chin[a] | TD=174 | .72 | .74 | .66 | .66 | .75 | .71 |
| 10-11 | Self/Hun[b] | TD=140 | .79 | .84 | .82 | .65 | .66 | .82 |
| 10-11 | Par/Hun[b] | | .86 | .89 | .86 | .71 | .61 | .76 |
| 10-11 | Tea/Hun[b] | | .96 | .94 | .91 | .81 | .76 | .88 |
| 10-12 | Par/Pers[c] | CP=230 | .76 | .74 | .61 | .62 | .68 | .72 |
| 10-12 | Self/Pers[d] | TD=114 | .69 | .78 | .67 | .67 | .68 | .63 |
| 10-12 | Par/Pers[d] | | .61 | .73 | .59 | .67 | .62 | .62 |
| 11-12 | Self/Chin[a] | TD=192 | .75 | .76 | .62 | .73 | .90 | .72 |
| 11-14 | Self/Chin[e] | TD=255 | .75 | .85 | .81 | .77 | .83 | .70 |
| 13-14 | Tea/Chin[e] | TD=66 | .94 | .93 | .90 | .92 | .90 | .56 |
| 11-18 | Self/Chin[f] | TD=239 | .70 | .87 | .85 | .78 | .87 | .59 |
| 16 | Self/Chin[g] | TD=235 | .79 | .88 | .83 | .85 | .88 | .75 |
| 8-15 | Par/Chin[h] | AD=64 | .80 | .86 | .85 | .77 | .85 | .79 |
| 11 | Self/Russ[i] | TD=167 | .82 | .90 | .85 | .85 | .83 | .77 |
| 11 | Self/Rom[i] | TD=150 | .85 | .91 | .82 | .80 | .79 | .79 |
| 8-18 | Par/Port[j] | TD=29 | .79 | .94 | .78 | .83 | .66 | .82 |

*Note. AD = ADHD; CP = cerebral palsy; Chin = Chinese; Hun = Hungarian; Negative reaction = Negative Reactions to Challenge; Par = Parent; Pers = Persian; Port = Portuguese; Rom = Romanian; Russ = Russian; Social w adults = Social Persistence with Adults; Social w children = Social Persistence with Children; Tea = Teacher; TD = typically developing.*

*[a] Huang (2019); [b] Józsa (2019); [c] Salavati et al. (2018); [d] Gharib et al. (2021); [e] Huang & Peng (2015); [f] Huang & Huang (2016); [g] Huang & Peng (2020); [h] Huang, et al. (2020); [i] Calchei et al. (2020); [j]Brandão et al. (2020)*

# Test-Retest Reliability

### Summary of Test-Retest Reliability for DMQ 17

Józsa and Molnár (2013) reported test-retest reliabilities, with a week to a month between ratings, ranging from .61 to .94 for 98 Hungarian teachers, parents, and school-aged students on the four instrumental and two expressive scales. The median correlations for these scales were .83, .80, and .74 for teacher, parents, and students, respectively. These test-retest correlations were highest for Object Oriented Persistence and Gross Motor Persistence, somewhat lower for the social persistence scales and Mastery Pleasure, and lowest for Negative Reactions to Failure. Miller et al. (2014) found good test-

retest reliabilities in their Australian sample for parent ratings of children with cerebral palsy; ICCs were .70 - .91 for the seven DMQ 17 scales.

### *Test-Retest Reliability for DMQ 18*

Table 4.4 provides test-retest reliabilities for 9 samples from 8 studies in 6 languages for 3-16 year-old children rated by themselves, a teacher, or a parent. Reliabilities of Hungarian preschool teachers' ratings and Bangladesh preschool teachers' ratings two weeks apart were acceptable to very good for all 6 DMQ 18 scales (Józsa & Morgan, 2015; Shaoli et al., 2019). Huang & Peng (2015) found acceptable but somewhat lower test-retest reliabilities from Taiwanese child self-ratings 1 month apart, except for the Negative Reactions to Challenge scale, which was unacceptable with a test-retest correlation of .54. Both Iranian typically developing schoolchildren and their parents and also parents of children with cerebral palsy had high (.70-.98) ICCs so good test-retest reliability for all scales given two weeks apart. These findings suggest that the lower alphas did not reflect general unreliability of the Persian version, but rather differences in how intercorrelated items from the same scale are. Hines and Bundy (2018) found acceptable ($r$ =.71) 10-day test-retest reliability for parent ratings of (only) cognitive persistence for Australian children with cerebral palsy. Also, Ramakrishnan (2015) found acceptable test-retest reliability of $r$ = .73 for homeless American parent ratings of these preschoolers' cognitive persistence.

Not shown in Table 4.4, the Competence scale test-retest reliabilities varied from .68 to .97 with a median of .85. Thus, there is good support for the test-retest reliability for the instrumental/persistence scales of DMQ 18 and acceptable to good test-retest reliability for all but one sample for the expressive/affective aspects of DMQ 18.

### *Stability Within a Developmental Stage for DMQ 18*

At this time, we do not have much stability data for DMQ 18. Hines and Bundy (2018) found strong 3- and 6-month stability (.76 and .76) for Australian ratings by a parent of their school-age child with cerebral palsy on the DMQ 18 Cognitive/Object Persistence scale. They did not report stability measures for the other DMQ scales.

Huang & Chen (2020) found good 6-8 month stability for Taiwanese ratings by parents of their children with developmental delay who ranged in age from 3 to 6 years (n = 40). Correlation coefficient were .72, .80, .56, .74, .64, and .68 for the six mastery motivation scales.

*Table 4.4. Test-Retest Reliability for DMQ 18 (ICC or Correlation Coefficients)*

| Age | Raters/ language | Child status /Ns | Instrumental/persistence aspect | | | | Expressive aspect | |
|---|---|---|---|---|---|---|---|---|
| | | | Cognitive/ object | Gross motor | Social w adults | Social w children | Mastery pleasure | Negative reactions |
| 3-6 yr | Teac/Hun[a] | TD=58 | .87 | .84 | .89 | .89 | .82 | .78 |
| 3-6 yr | Teac/Ban[b] | TD=50 | .84 | .88 | .86 | .88 | .79 | .89 |
| 5-8 yr | Teac/Kis[c] | TD=30 | .80 | .89 | .82 | .86 | .94 | .89 |
| 10-12 yr | Self/Pers[d] | TD=33 | .91 | .89 | .93 | .95 | .94 | .97 |
| 11-14 yr | Self/Chin[e] | TD=251 | .71 | .73 | .70 | .70 | .69 | .54 |
| 10-12 yr | Par/Pers[f] | CP=32 | .91 | .85 | .96 | .79 | .84 | .84 |
| 10-12 yr | Par/Pers[d] | TD=42 | .85 | .89 | .79 | .85 | .72 | .77 |
| 8-16 yr | Par/Eng[g] | CP=19 | .71 | N/A | N/A | N/A | N/A | N/A |
| 3-6 yr | Par/Eng[h] | HL=36 | .73 | N/A | N/A | N/A | N/A | N/A |

*Note. Ban = Bangla; Chin = Chinese; CP = Cerebral palsy; DD = Developmental Delay; Eng = English; HL = Homeless; Hun = Hungarian; Kis = Kiswahili; NA = not available; Par = Parent; Pers = Persian; Teac = Teacher; TD = Typically Developing.*
*[a] Józsa & Morgan (2015); [b] Shaoli et al. (2019); [c] Amukune et al. (2020), a few of these pre-school children in Kenya were as old at 12 years, but 52% were 5-6 and 86% were 5-8; [d] Gharab et al. (2020); [e] Huang & Peng (2015); [f] Salavati et al. (2018); [g] Hines & Bundy (2018); [h] Ramakrishnan (2015).*

# Interrater Reliability

### *Summary of Interrater Reliability for DMQ 17*

An analysis of Hungarian DMQ 17 data was carried out by examining the correlations between the ratings of pairs of teachers who rated the same children but in somewhat different contexts (Józsa & Molnár, 2013). One of the teacher raters was the homeroom teacher and the other was a teacher who taught the children in several courses. Correlations between the ratings of total mastery motivation by these teachers for children in grades 4 and 8 were moderate, indicating a relatively close correspondence between teacher ratings. However, in grade 10, much lower correlations were found. This may be because in grade 10, the teachers teach the children in only one subject (e.g. math or history) so they know the children in different contexts and less well than the teachers in grades 4 and 8.

### *Interrater Reliability for DMQ 18*

Table 4.5 shows that interrater reliabilities for Hungarian preschool teachers were minimally adequate to very good using intraclass correlation coefficients (ICC) based on ratings of preschool children by each of the child's two teachers (Józsa & Morgan, 2015). Except for Gross Motor Persistence, there was acceptable to good interrater reliability on each of the persistence scales and Mastery Pleasure. However, the alpha for Negative Reactions to Challenge was only minimally acceptable and was inadequate for the two negative reactions subscales. The alpha was .87 for Competence. Apparently, the child's two preschool teachers see Gross Motor Persistence and Negative Reactions to Challenge differently, but have little trouble evaluating and agreeing on a child's ability or competence and their cognitive persistence relative and to other children.

In the Bangladesh sample, the correlations between Bangla-speaking teacher and parent ratings were high, indicating very good interrater reliability.

**Table 4.5. Interrater Reliability for DMQ 18**

| Age | Raters/ language | Child status/ Ns | Instrumental/persistence aspect | | | | Expressive aspect | |
|---|---|---|---|---|---|---|---|---|
| | | | Cognitive/ object | Gross motor | Social w adults | Social w children | Mastery pleasure | Negative reactions |
| **3-6 yr** | T, T/Hun[a] | TD=133 | .85 | .65 | .78 | .79 | .78 | .61 |
| **3-6 yr** | T, P/Ban[b] | TD=30 | .85 | .86 | .80 | .83 | .88 | .85 |

*Note. Ban = Bangla; Hun = Hungarian; P = Parent; T = Teachers; TD = Typically developing.*
[a]*Józsa & Morgan (2015);* [b]*Shaoli et al. (2019)*

# Parallel Forms Reliability

### *Summary of Parallel Forms Reliability for Earlier DMQ Versions*

The DMQ-G items were modified, mostly in minor ways, to make the DMQ easier to answer. The equivalence of the DMQ-G general scale scores with the revised and expanded DMQ-E was tested by asking mothers of 35 children, 29- to 59-months old, to answer both versions about three weeks apart. Half answered the revised version first, and half answered it second. These correlations (general persistence, .85; overall mastery pleasure, .70; independent mastery attempts, .83; and general competence, .58) indicated that the scale scores of the two versions were quite highly related. For general persistence, the correlations indicated good alternate forms reliability. The overall correlation for mastery pleasure was acceptable but somewhat lower because we attempted to differentiate two related but somewhat distinct concepts: pleasure during the process of goal-directed behavior and

pleasure at causing something to happen. As expected, the correlation was somewhat lower for competence because several items had been changed to improve the psychometric properties of the scale and to try to differentiate competence more clearly from persistence.

### *Parallel Forms Reliability for DMQ 18*

Józsa and Morgan (2015) asked the same teachers to rate using both DMQ 17 and DMQ 18. These were not really parallel forms because a number of items were deleted and others were changed from DMQ 17 to DMQ 18, as noted in **Chapter 2**. However, these two versions of the DMQ had the same scales and many of the same items, so the correlations in Table 4.6 are similar to parallel forms reliability coefficients. Note that the negative reactions items were changed dramatically, which accounts for the relatively low correlation.

*Table 4.6. Correlations to Assess Parallel Forms Reliability of DMQ 18*

| Age | Raters/ language | Child status/ Ns | Instrumental/persistence aspect | | | | Expressive aspect | |
|---|---|---|---|---|---|---|---|---|
| | | | Cognitive/ object | Gross motor | Social w adults | Social w children | Mastery pleasure | Negative reactions |
| **3-6 yr** | T17-T18/ Hun[a] | TD=30 | .63 | .60 | .76 | .65 | .59 | .38 |
| **3-6 yr** | T/Eng – T/Ban[b] | TD=20 | .87 | .86 | .74 | .85 | .78 | .72 |
| **5-8 yr** | T/Eng – T/Kis[c] | TD=20 | .80 | .57 | .87 | .82 | .76 | .73 |

*Note. Ban = Bangladesh; Eng = English; Hun = Hungarian; Kis = Kiswahili; T = Teacher rating; TD= typical development.*
*[a]Józsa and Morgan (2015); [b]Shaoli et al. (2019); [c]Amukune et al. (2020), a few of these preschool children in Kenya were as old at 12 years, but 52% were 5-6 and 86% were 5-8.*

Shaoli et al. (2019) examined the correlations between the same teachers' ratings of the English and the Bangla version of DMQ 18 (see Table 4.6). The correlations were quite high, ranging from .72-.87, providing both evidence that DMQ measures similar constructs in the two languages and that teacher ratings were reliable.

Similarly, Amukune et al. (2020) correlated the English and Kiswahili versions of the preschool DMQ 18 rated by the same Kenyan teachers. These ratings were again acceptable for all the scales, including Negative Reactions to Challenge, the scales except Gross Motor Persistence.

## Conclusion

This chapter presented evidence for a number of ways of assessing evidence for the reliability of the DMQ 18 in 12 languages with 33 samples of infant,

preschool, and school-age children, both children developing typically and atypically. The bulk of the evidence is supportive of the reliability of the DMQ 18 data, as was the evidence for the reliability of DMQ 17. As discussed in **Chapter 2,** DMQ 18 was carefully developed by researchers in the US, Taiwan, and Hungary using statistical analyses of DMQ 17 data and the process of decentering in order to make the questionnaire more appropriate to translate and adapt to other cultures.

It is not possible to compare directly alphas for DMQ 17 and DMQ 18 because a number of items were deleted or revised and because the DMQ 18 reliability data come from nine new languages in addition to the three used to develop it. The DMQ 18 reliability data were based on smaller samples of a larger set of languages, often for the first study using that language version of the DMQ. Nevertheless, reliability measures for DMQ 17 and 18 are similar. Alphas were acceptable for the persistence scales and Mastery Pleasure, and DMQ 18 had somewhat better reliabilities for overall Negative Reactions to Challenge.

In terms of **internal consistency** (i.e., Cronbach alphas) for DMQ 18, 90% of the four persistence scales for infants and preschool children had acceptable alphas ($\geq$ .70) and the remaining 10% were minimally acceptable. For Mastery Pleasure and overall Negative Reactions to Challenge, 94% of the scales had acceptable alphas for infants and preschool children, and all the rest were minimally acceptable.

For 8-18 year-old school children, 95% of the internal consistency alphas for the persistence scales were acceptable for the Chinese, Hungarian, Russian, Romanian, and Portuguese-speaking samples. The Iranian Persian-speaking samples were more problematic for both the persistence scales and the expressive/affective scales, with most being marginally acceptable, and only 1 of 18 being unacceptable. For the non-Iranian samples, all of the Mastery Pleasure alphas were acceptable, with only three being marginally acceptable. However, two of the non-Iranian Negative Reactions to Challenge alphas were unacceptable.

There did not seem to be any clear differences in alphas for children developing typically and children at risk or developing atypically. There also did not seem to be clear differences between the alphas for the different languages, except for somewhat lower alphas for the school-age Persian-speaking children, which were almost all at least minimally acceptable.

**Test-retest reliabilities** were adequate to very good for all of the persistence scales in all six languages that reported this type of data. Only one sample out of seven had a minimally acceptable coefficient for Mastery Pleasure, and a school age sample had an unacceptable test-retest reliability for Negative Reactions to Challenge.

**Interrater reliability** was at least minimally acceptable for the two DMQ 18 studies that reported this type of data, which is difficult to obtain

because it is unusual for any two raters (e.g., parent and teacher) to see the same child in the same context

Again, because there is only one version or form of DMQ 18, we can only approximate **parallel forms reliability.** One study correlated DMQ 17 and 18 scale scores and reported significant correlations between them, except for negative reactions, whose items had been changed a lot. Two other studies asked the same raters to rate the DMQ in English and in the native language and reported significant and mostly high correlations.

In conclusion, all the measures of reliability provided evidence to support the reliability of the DMQ 18 data in 12 different languages and for infants, preschool, and school-age children, both those developing typically and those developing atypically.

The next chapter summarizes the evidence for measurement validity of the DMQ, using evidence from both DMQ 17 and DMQ 18.

# References

Amukune, S., Calchei, M., & Józsa, K. (2020). *Adaptation of the Preschool Dimensions of Mastery Questionnaire (DMQ 18) for preschool children in Kenya.* [Manuscript submitted for publication].

Blasco, P. M., Acar, S., Guy, S., Saxton, S. N., Dasgupta, M., & Morgan, G. A. (2020). Executive function in infants born low birth weight and preterm. *Journal of Early Intervention. 42,* 321–327, https://doi.org/10.1177/1053815120921946

Blasco, P. M., Gerton, J. M., Acar, S., Guy, S., & Saxton, S. (2019). *Unpublished DMQ 18 internal consistency reliability data from parents of 18-month-old children who are preterm or full-term.* Western Oregon University.

Brandão, M., Mancini, M., C., Figuieredo, P., Oliverira, R., Avelar, B. (2020). *Unpublished DMQ 18 reliability data from parent ratings of infants, preschool, and school-age children.* Universidade Federal de Minas Gerais, Brazil.

Calchei, M., Amukune, S., & Józsa, K. (2020). *Adaptation of Dimensions of Mastery Questionnaire (DMQ 18) for school-age children into the Russian and Romanian languages.* [Manuscript submitted for publication].

Chang, C.-Y., Huang, S.-Y., & Tang, S.-C. (2020). *Analyses of DMQ 18 pre-school version-Chinese for Taiwan toddlers with suspected speech delay* [Unpublished analyses]. Fu Jen Catholic University, Taipei, Taiwan.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. https://doi.org/10.1007/BF02310555

Gharib, M., Vameghi, R., Hosseini, S. A., Rashedi, V., Siamian, H., & Morgan, G. A. (2021). Mastery motivation in Iranian parents and their children: A comparison study of their views. *Journal of Nursing and Midwifery Sciences, 8*(1), 54–60.

Gilmore, L., & Boulton-Lewis, G. (2009). Just try harder and you will shine: A study of 20 lazy children. *Australian Journal of Guidance and Counseling, 19(2),* 95–103.

Hines, A., & Bundy, A. (2018). *Unpublished DMQ 18 data for parent's ratings of school-aged children with cerebral palsy.* University of Sydney, Australia.

Huang, S.-Y. (2019). *Reliability of self-rated school-age DMQ 18 scales and correlations of DMQ 18 with school achievement in 5th and 6th grade children in Taiwan* [Unpublished analysis]. Fu Jen Catholic University, Taipei, Taiwan.

Huang, S.-Y., & Chen, H.-W. (2020). *Analyses of DMQ 18 preschool version-Chinese for Taiwan children with developmental delay* [Unpublished analyses]. Fu Jen Catholic University, Taipei, Taiwan.

Huang, K.- T., & Huang, S.- Y. (2016). *The reliability and validity of DMQ 18 school version-Chinese.* Poster at the conference of Chinese psychology, Taiwan, Tainan.

Huang, S.-Y., & Lo, P. (2019). *Reliability of toddler and preschool DMQ 18 and correlations with BSID-III or WPPSI-IV IQ* [Unpublished analyses]. Fu Jen Catholic University, Taipei, Taiwan.

Huang, S.-Y., & Peng, Y.-Y. (2015). *Analyses of DMQ 18 and subject specific mastery motivation data for 5th to 8th grade Taiwan students and teachers* [Unpublished analyses]. Fu Jen Catholic University, Taipei, Taiwan.

Huang, S.-Y., & Peng, Y.-Y. (2020). *Analyses of DMQ 18 school version-Chinese for 10th grade Taiwan students* [Unpublished analyses]. Fu Jen Catholic University, Taipei, Taiwan.

Huang, S.-Y., Zang, X-H, & Tsai., F.-J. (2020). *Unpublished analyses of DMQ 18 school version-Chinese for Taiwan school-age children and adolescents with ADHD.* Fu Jen Catholic University, Taipei, Taiwan.

Józsa, K. (2019). *Reliability and validity of DMQ 18 of 4th grade students in Hungary rated by themselves, parents, and a teacher* [Unpublished data]. University of Szeged, Hungary.

Józsa, K., & Barrett, K. C. (2018). Affective and social mastery motivation in preschool as predictors of early school success: A longitudinal study. *Early Childhood Research Quarterly, 45*(4), 81–92.

Józsa, K., & Molnár, É. D. (2013). The relationship between mastery motivation, self-regulated learning and school success: A Hungarian and European perspective. In K. C. Barrett, N. A. Fox, G. A. Morgan, D. J. Fidler, & L. A. Daunhauer (Eds.), *Handbook on self-regulatory processes in development: New directions and international perspectives* (pp. 265–304). Psychology Press.

Józsa, K., & Morgan, G. A. (2015). An improved measure of mastery motivation; Reliability and validity of the Dimensions of Mastery Questionnaire (DMQ 18) for preschool children. *Hungarian Educational Research Journal, 5*(4), 87–103. https://doi.org/10.14413/HERJ2015.04.08

Józsa, K., Wang, J., Barrett, K. C., & Morgan, G. A. (2014). Age and cultural differences in mastery motivation in American, Chinese, and Hungarian school-age children. *Child Development Research, 2014,* Article ID 803061, 1–16. https://doi.org/10.1155/2014/803061

Miller, L., Marnane, K., Ziviani, J., & Boyd, R. N. (2014). The Dimensions of Mastery Questionnaire in school-aged children with congenital hemiplegia: Test-retest reproducibility and parent-child concordance. *Physical & Occupational Therapy in Pediatrics, 34*(2), 168–184.

Morgan, G. A., Liao, H.-F., Nyitrai, A., Huang, S.-Y., Wang, P.-J., Blasco, P., Ramakrishnan, J., & Józsa, K. (2017). The revised Dimensions of Mastery Questionnaire (DMQ 18) for infants and preschool children with and without risks or delays in Hungary, Taiwan, and the US. *Hungarian Educational Research Journal, 7*(2), 48–67. https://doi.org/10.14413/HERJ/7/2/4

Morgan, G. A., Wang, J., Liao, H.-F, & Xu, Q. (2013). Using the Dimensions of Mastery Questionnaire to assess mastery motivation of English- and Chinese- speaking children: Psychometrics and implications for self- regulation. In K. C. Barrett, N. A. Fox, G. A. Morgan, D. J. Fidler, & L. A. Daunhauer. (Eds.), *Handbook of self-regulatory processes in development: New directions and international perspectives* (pp. 305–335). Psychology Press.

Özbey, S. (2020). *Means, SDs and Cronbach's alphas from 1592 Turkish preschool children rated by teachers.* [Unpublished data analyses] Gazi University, Ankara, Turkey.

Rahmawati, A., Fajrianthi, Morgan, G. A., & Józsa, K. (2020). An adaptation of DMQ 18 for measuring mastery motivation in early childhood. *Pedagogika*, *140*(4), 18–33.

Ramakrishnan, J. (2015). *DMQ 18 data from mothers of 3-5 year-old children living at a Minneapolis emergency homeless shelter* [Unpublished data]. Institute of Child Development, University of Minnesota.

Salavati, M., Vameghi, R., Hosseini, S. A., Saeedi, A., & Gharib, M. (2018). Mastery motivation in children with cerebral palsy (CP) based on parental report: Validity and reliability of Dimensions of Mastery Questionnaire in Persian. *Materia Socio Medica, 30*(2), 108–112.

Saxton, S. N. Blasco, P. M., Gullion, L., Gerton, J. M., Atkins, K., & Morgan, G. A. (2020). *Examination of mastery motivation in children at high risk for developmental disabilities*. [Manuscript submitted for publication].

Shaoli, S. S., Islam, S., Haque, S., & Islam, A. (2019). Validating the Bangla version of the Dimensions of Mastery Questionnaire (DMQ 18) for preschoolers. *Asian Journal of Psychiatry, 44*, 143–149.

Wang, P.-J., & Lewis, A. (2019). *DMQ 18 Data from a study of preschool children in Colorado developing typically* [Unpublished data]. Colorado State University. Fort Collins, CO.