The psychometric properties of a brief version of the Systemizing Quotient

Jaimie F. Veale[a]

Matt N. Williams[b]

[a]Department of Psychology, University of British Columbia, Vancouver, Canada.

[b]Corresponding author. School of Psychology, Massey University, Auckland, New Zealand.

m.n.williams@massey.ac.nz

Veale, J. F., & Williams, M. N. (2015). The psychometric properties of a brief version of the Systemizing Quotient. *European Journal of Psychological Assessment.* Advance online publication. http://dx.doi.org/10.1027/1015-5759/a000283

Summary

The construct of systemizing—the drive to construct or understand systems—has an important role in the Extreme Male Brain theory of autism. While a brief version of the Systemizing Quotient (SQ) has been proposed, there is a need to assess its psychometric properties. This study assessed factorial and construct validity of an 8-item version of the SQ on a sample of 627 participants. A single-factor latent variable model with a single correlated error term showed adequate fit in a confirmatory factor analysis. This model also demonstrated metric invariance across genders when controlling for an effect of age on item responses. Reliability was acceptable, $\alpha = .72$. As further evidence for construct validity, SQ scores showed expected relationships with mental rotation performance, study area, trait anxiety, childhood extroversion, childhood agreeableness, and gender. Overall, the results indicated good psychometric properties for the brief version of the SQ, suggesting that this scale could be useful when researchers require a systemizing measure that is minimally burdensome to complete.

Keywords: systemizing, Systemizing Quotient, confirmatory factor analysis, invariance testing.

Introduction

The concept of systemizing has an important role in the Extreme Male Brain Theory of Autism (Baron-Cohen, 2002). Systemizing refers to a person's propensity to understand and construct systems. Examples of systems include computers, musical instruments, weather, mathematics, political systems, and library organizing systems. Baron-Cohen proposed that people with autism have markedly greater systemizing ability (a trait more common in males) than their empathizing ability (a trait more common in females). As the name of the theory suggests, those scoring higher on systemizing are hypothesized to have more of a "male brain."

The systemizing quotient (SQ) was developed to measure the construct of systemizing. The original SQ had 40 items that assessed systemizing (Baron-Cohen, Richler, Bisarya, Gurunathan, & Wheelwright, 2003). A revised version (the SQ-R) was subsequently developed to incorporate items that "might be more relevant to females in the general population…to test if systemizing scores are higher among males even with the inclusion of items selected from traditionally female domains" (Wheelwright et al., 2006, p. 49). Shortened versions of the SQ have also been proposed based on analyses of the original 40-item scale. Using principal component analysis, Wakabayashi et al. (2006) proposed a 25-item one-factor scale that they called the short-form. From a confirmatory factor analysis, Ling, Burton, Salt, and Muncer (2009) found that an 18-item scale with four subfactors had the best fit. The factors were labeled Technicity, Topography, Do It Yourself (DIY), and Structure.

In studies that measure a large number of constructs, participant response burden may be a concern, potentially causing reduced response rates, reduced survey completion, and lower data quality. A measure of systemizing that is even shorter than 18 items might be preferable to researchers in such cases. Indeed, researchers in a recent large online survey of

sex differences across a variety of domains selected the 10 items from the SQ-R that most differentiated sex and called this the "brief SQ questionnaire" (Manning, Baron-Cohen, Wheelwright, & Fink, 2010, p. 769). Manning and colleagues did not assess the psychometric properties of this brief version of the questionnaire. Such an assessment is needed to potentially allow researchers to assess systemizing using a validated scale that is minimally burdensome.

This study uses confirmatory factor analysis to assess the factorial validity and factor structure of responses to the brief SQ items. Although Ling and colleagues (2009) found a four-factor structure, the items in Manning et al.'s (2010) brief version of the SQ does not provide adequate coverage of all of these four factors. Furthermore, the intent of the scale is to measure an overarching single systemizing factor, so this study tested a unidimensional factor model.

Measurement invariance testing can be conducted to assess whether the factor structure (factor loadings and intercepts) are equivalent (invariant) across groups. This is important because comparisons of scales across groups may be invalidated if the factor being measured is not invariant across the groups. For example, estimates of the difference between genders in mean SQ scores are not fully meaningful if the factor structure differs between genders. This can occur if there are differences between the groups in the perception or interpretation of the scale items. Measurement invariance testing has also been called differential item functioning testing among those from the item response theory tradition. For more information about invariance testing, see Gregorich (2006). No research has previously assessed the equivalence of the factorial structure of any version of the SQ across groups. To address this gap, we conducted invariance testing across genders and across countries.

Construct validity of the brief version of the SQ was further assessed by testing whether scores on the test showed expected relationships with a number of other variables. In

accordance with theory (Baron-Cohen, 2002) and prior research (Baron-Cohen et al., 2003; Nettle, 2007; Wakabayashi et al., 2006; Ling et al., 2009), we hypothesized that males would have higher brief SQ scores than females. Baron-Cohen (2002) noted that on average males outperform females on the mental rotation test and he proposed that this is due to the systemic nature of the task, with predictions of how the rotated figures would look as the systemic output. Previous studies have found a positive relationship between SQ scores and performance on a mental rotation task (Cook & Saucier, 2010; Ling et al., 2009), so we hypothesized that we would also see this correlation with the brief SQ. We also expected a small negative relationship between systemizing and trait anxiety as this was found in one previous study (Pingault, Pouga, Grèzes, & Berthoz, 2012). In an examination of discriminant validity, tests of childhood extroversion and agreeableness were expected to have very small or zero relationships with systemizing (per the findings of Nettle, 2007).

Methods

Participants

Participants were recruited for an internet-based survey described as investigating the development of gender and sexuality. This was conducted through Google advertising, contacting GLBTQ-related organizations and online groups, and via a press release through Massey University Communications, generating media attention. This research was conducted as part of a larger survey study that assessed biological and psychosocial correlates of gender identities (Veale, Clarke, & Lomax, 2010) and included a number of transgender participants. In order to generalize to a wide population, we excluded transgender participants from the analysis, so the 627 participants described here were all not transgender. Only participants of the larger study who had completed at least one item of the brief SQ were included. A total of 379 (60%) participants were male and 248 were female (40%). Countries of residence were the USA (51%), New Zealand (28%), Great Britain (8%),

Canada (3%), Australia (2%), and 7% were from other countries. The most common reported ethnicities were European (93%), Asian (5%), and Hispanic (2%). The participants were generally quite highly educated: 55% had a Bachelor's degree or higher qualification, and only 5% had completed three years or less of high school. The mean age was 38 years (*SD* = 14 years).

Measures

Systemizing.

The ten items of the brief SQ questionnaire created by Manning and colleagues (2010) were considered for inclusion in the study. Two items were removed: one asking about respondents' enjoyment of political discussions and one asking about their knowledge of legislative process. We believed that these items would not be appropriate for an international sample because we expected that the questions would have different meanings across countries with different legislative processes, cultural acceptability of political discussions, and freedom for political discussions. This left eight items in the scale, which are displayed in Table 1.

Items used a 7-point Likert response scale: *strongly agree*/*disagree*, *moderately agree*/*disagree*, *slightly agree*/*disagree*, and *neither agree or disagree*. The response scale wording differed slightly from previous versions of the questionnaire which used four response options: *definitely agree/disagree* and *slightly agree/disagree*. A midpoint response option and additional three response options were added. A larger number of response options has been shown to be advantageous to scale reliability and validity, but with diminishing gains if more than 7 options are given (Lozano, García-Cueto, & Muñiz, 2008; see also Preston & Colman, 2000).

Mental rotation.

In an online adaptation of Vandenberg and Kuse's (1978) mental rotation test, participants viewed two three-dimensional cuboids and decided whether they were the same (only rotated) or different objects. Participants had two minutes to complete up to 34 items. They gained one point for every correct answer but lost one point for every incorrect answer. Unattempted items received zero points. We excluded 147 participants who did not attempt any items from analysis that used this test.

Discriminant validity items.

Each participant was asked five items pertaining to their agreeableness as a child, and ten relating to their level of extroversion as a child using the 7-point Likert scale International Personality Item Pool questions (Goldberg et al., 2006). Trait anxiety was assessed using the 9-item version of the Taylor Manifest Anxiety Scale (Taylor, 1953) with response on a 7-point Likert scale.

Procedure

Approval for the study was obtained by Massey University Human Research Ethics Committee. Participants consented to the study by completing the questionnaire.

Data analysis.

One percent of the total possible item responses for the brief SQ scale were missing. Across the other personality measures (those of childhood extroversion and agreeableness, and trait anxiety), 16.2% of possible item responses were missing. Confirmatory factor analysis models were estimated using full information maximum likelihood, which allows for the presence of missing data, with robust standard errors and test statistics scaled for non-normality (MLR estimation). The remaining analyses were conducted using conventional methods which do not allow the presence of missing data, so for these analyses missing data were multiply imputed (5 imputed datasets) using predictive mean matching in the R package

mice (Buuren & Groothuis-Oudshoorn, 2011), and analyses pooled across imputed datasets per Rubin (1987).

Total scores on the SQ were obtained by summing the responses to the eight items, wherein at the item level a *strongly disagree* response was coded as 0, *moderately disagree* was coded as 1, and so forth (after reverse-coding negatively worded items). This yielded a possible score range of 0–48.

Data analyses were completed in R version 3.0.2. Confirmatory factor analysis was completed using the package lavaan version 0.5–14 (Rosseel, 2013), with invariance testing completed using the semTools package version 0.4–0 (Pornprasertmanit et al., 2013).

Invariance testing involved fitting a sequence of four models: a configural invariance model (the same model fit in both groups, but with parameters free to vary across groups); a weak invariance model (factor loadings held constant across groups); a strong invariance model (both factor loadings and intercepts held constant across groups); and a model with latent means held equal across groups.

Results

Descriptive Statistics

Descriptive statistics for the brief SQ items are displayed in Table 1. Variation in responses was smallest for items 1 and 5, with most participants scoring highly on these items. Skewness and kurtosis was also greatest for these two items, although in general item distributions differed from a normal distribution—as will always be the case for items with only a discrete set of response options.

Table 1

Descriptive statistics for brief SQ items used in this study

Item	Full wording	Mean	Median	SD	Skewness	Kurtosis
1	I find it difficult to read and understand maps*	5.0	6	1.6	-1.6	4.2
2	I find it easy to grasp exactly how odds work in betting	3.1	3	2.1	0.0	1.6
3	I find it difficult to learn how to programme video recorders*	4.3	5	2.0	-0.9	2.5
4	I do not enjoy games that involve a high degree of strategy (e.g. chess, Risk, Games Workshop)*	3.5	4	2.1	-0.4	1.7
5	I can remember large amounts of information about a topic that interests me e.g. flags of the world, airline logos	4.7	5	1.6	-1.5	4.3
6	I can easily visualise how the motorways in my region link up	4.1	5	2.0	-0.8	2.3
7	I am fascinated by how machines work	3.4	4	2.1	-0.3	1.7
8	If I were buying a stereo, I would want to know about its precise technical features	3.1	4	2.2	-0.1	1.5

Note. All items had a possible score range of 0–6. *Item is scored such that higher levels of agreement indicates lower levels of systemizing.

Reliability

Cronbach's alpha for the brief SQ was .72, 95% confidence interval [.70, .74]. Analysis of individual items indicated that deletion of item 5 would improve alpha very

slightly to .73. The alpha value did not differ meaningfully by gender: $\alpha = .71$ for males, and $\alpha = .73$ for females.

Confirmatory Factor Analysis

A model with all items loading on a single factor was tested. For this one-factor model, the chi-square fit statistic indicated that a null hypothesis that the model exactly explains the correlations between items in the population could be rejected, $\chi^2(20) = 131.36$, $p < .001$. The root mean squared error of approximation (RMSEA) of .094 also indicated "mediocre" fit of the model relative to its complexity, per the guidelines of MacCullum, Browne, and Sugawara (1996, p. 134), and the comparative fit index (CFI) of .846 fell beneath the .95 cutoff for good fit suggested by Hu and Bentler (1999). On the other hand, the standardized root mean squared residual (SRMR) of .051 suggested good fit according to the SRMR < .08 cutoff advocated by Hu and Bentler.

A large modification index of 69.1 suggested that allowing the error terms for item 1 ("I find it difficult to read and understand maps") and item 6 ("I can easily visualise how the motorways in my region link up") to correlate would markedly improve model fit. The content in these two items is very closely related and it seems reasonable to assume that there would be a larger correlation between responses to these items than can be explained purely by a systemizing factor common to all items. An alternative model with correlated errors between these items was therefore estimated and this is depicted in Figure 1. For this model, the chi-square statistic still indicated that an hypothesis of exact fit could be rejected $\chi^2(19) = 71.69$, $p < .001$ and the CFI of .927 fell just below the cutoff of close to .95 for good fit suggested by Hu and Bentler (1999). However, the RMSEA of .067 now fell within the "reasonable" fit range suggested by Browne and Cudeck (1992, p. 239), and the SRMR of .039 was suggestive of a small magnitude of model error. The overall fit of the one-factor

model with a correlated error term was therefore imperfect but reasonably adequate, and substantially improved from the previous model.

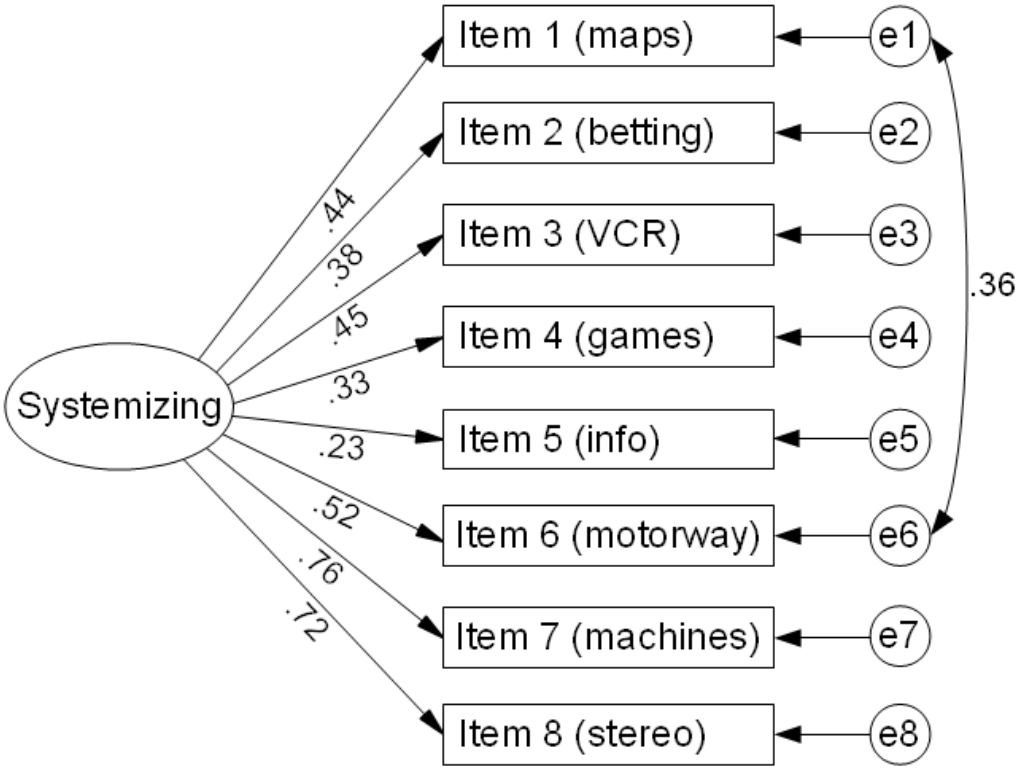

Figure 1. The tested factor model with completely standardized parameter estimates. All coefficients shown have $p < .001$.

Invariance Testing

We assessed invariance of the modified single factor model across males and females (see Table 2). A configural invariance model had reasonable fit, RMSEA = .066. Constraining factor loadings to be equal across groups resulted in no significant loss of fit and an improvement in the parsimony-adjusted Bayesian Information Criterion (BIC). However, holding intercepts equal across groups resulted in a significant loss of fit, and an

increase (worsening) of the BIC statistic. Modification indices suggested that this noninvariance seemed to relate especially to the intercept for item 3 ("I find it difficult to learn how to programme video recorders"). Males seemed to report more difficulty with video recorders than women, independently of systemizing level. We suspected this was because the male sample had a higher average age, $M = 42$ years, than the female sample, $M = 30$ years, which likely affected their responses to items relating to use of technology (also including items 7 and 8). Indeed, in an alternative exploratory analysis of invariance across older and younger participants (via a median split), there was a substantial loss of fit if intercepts were held equal across age groups, with a large increase in the chi-square statistic in comparison to a weak invariance model, $\Delta\chi^2(7) = 78.01$, $p = < .001$. (Further information about this test of invariance across age groups can be found in the Electronic Supplementary Materials).

We therefore tested invariance across gender of the model reported previously, but including age as an observed predictor of responses to each of the systemizing items. Thus, responses to brief SQ items were assumed to be produced by linear effects of both systemizing level and age in this model. This modification resulted in reasonable fit for the configural invariance model, $\chi^2(38) = 90.98$, RMSEA = .067, BIC = 25,043, with no loss of fit if both loadings and intercepts were held constant in a strong invariance model, $\Delta\chi^2(14) = 10.89$, $p = .695$, BIC = 24,965, RMSEA = 0.056. As such, when controlling for the different mean ages of the male and female subsamples, there was reasonable evidence of strong invariance of the scale across gender.

Table 2

Invariance testing across genders

Model	χ^2	df	*p*	CFI	BIC	$\Delta\chi^2$	Δdf	*p*	ΔCFI
Unconstrained	90.55	38	< .001	.925	20,092	-	-	-	-
Factor loadings equal	97.41	45	< .001	.925	20,054	6.68	7	.463	.000
Loadings and intercepts equal	149.84	52	< .001	.861	20,066	59.94	14	<.001	.065
Loadings, intercepts, and means equal	176.44	53	< .001	.824	20,088	87.71	15	< .001	.101

Note. Statistics marked with Δ indicate the change in fit in comparison to unconstrained model. Male $n = 379$, female $n = 248$

Holding latent means equal across genders resulted in a loss of fit in comparison to the strong invariance model, $\Delta\chi^2(1) = 5.08$, $p = .024$. This suggests that men and women differed in their levels of the Systemizing factor. In the model with equal factor loadings and intercepts, and age controlled, males had higher Systemizing levels than females, latent mean difference = 0.44, Cohen's $d = 0.64$. A similar gender difference was found when analyzing simple brief SQ sum scores, $M = 32.5$, $SD = 9.0$ for males, and $M = 28.9$, $SD = 9.2$ for females, $p < .001$, albeit with a smaller effect size, $d = 0.39$.

Invariance testing between North Americans ($n = 340$) and Australasians ($n = 191$) was also conducted (see Electronic Supplementary Materials). This analysis found factor loading invariance and intercept invariance when controlling for age.

Construct Validity

The invariance analyses presented above support one of our key construct validity predictions: that males would exhibit higher systemizing levels than females. The results of

other analyses relating construct validity tests are summarized in Table 3, with the findings largely being in accordance with predictions.

Table 3

Summary of construct validity predictions and results

Prediction	Result	Prediction supported?
Positive correlation between SQ and mental rotation scores	$r(n = 480) = .24$*	Yes
Negative correlation between SQ and trait anxiety	$r(N = 627) = -.23$*	Yes
Correlation between SQ and childhood extroversion near zero	$r(N = 627) = -.04$	Yes
Correlation between SQ and agreeableness near zero	$r(N = 627) = -.10$	Yes

Note. * $p < .05$. Correlations pooled across multiply imputed datasets using Fisher transformations, with standard errors calculated per Rubin (1987).

To further investigate discriminant validity, the items of the brief SQ were entered into an exploratory factor analysis along with all other rating scale items used in this study (those assessing childhood extroversion, childhood agreeableness, and trait anxiety). Principal axis factoring with direct oblimin rotation was used. Velicer's (1976) MAP criterion suggested a 4-factor solution, which was also the number of scales analyzed. For this 4-factor model, all of the brief SQ items had their largest loading on the same factor, with the smallest loading being .24 for item 5 (information). No item from any other scale had a substantial positive loading (i.e. > .1) on this Systemizing factor.

Discussion

The results of our study generally supported the reliability and validity of the brief version of the SQ. Confirmatory factor analysis indicated that a single factor model with correlated errors between two items both relating to topography showed adequate fit. This suggests that the brief version of the SQ has adequate psychometric properties and may be useful for researchers wanting a brief measure of the systemizing construct.

While a previous study by Ling et al. (2009) found support for a four-factor model, it was not feasible to test a 4-factor model in the current study because the short length of the scale meant that there was very limited coverage of these four factors. Furthermore, the intent of the scale is to measure an overarching single systemizing factor, so this study tested a unidimensional model. In accordance with our findings, another study using an 18-item version of the systemizing quotient found acceptable fit for a one-factor model (Morsanyi, Primi, Handley, Chiesi, & Galli, 2012). Ultimately, the most appropriate model may be a higher-order model incorporating both four lower-order factors as well as an overarching systemizing factor. Ling and colleagues did not assess such a model and we suggest that this be tested in future research using longer versions of the scale.

Although some of the key claims of the empathizing-systemizing theory are made with respect to the differences in the levels of these constructs across genders, this was the first study to assess measurement invariance of a systemizing questionnaire across genders. Our results showed measurement invariance for factor loadings across males and females, providing evidence that the items on the brief version of the SQ have similar relationships with the systemizing factor across genders. While there was a gender difference in intercepts, further analyses showed that this was likely to be due to age differences between the gender groups in our sample. We found a tendency for older participants to report more difficulty with the use of technology, independent of their level of systemizing. After controlling for

the effect of age on item responses, there was positive evidence for strong invariance (i.e., invariance of both factor loadings and intercepts) of the scale across genders. This was despite the fact that the items in brief SQ were initially selected due to the presence of large sex differences on these items (Manning et al., 2010). This means that we can conclude for the first time that that the sex difference in systemizing scores are in fact due to a difference in systemizing levels, as opposed to differences between genders in the measurement of the construct. This study also found invariance across North American and Australasian country groups once age was controlled for, but slight adjustments to terminology may be required to suit the linguistic conventions of particular populations (e.g. changing "motorways" to "freeways").

However, this research also found evidence of a problem that the scale is *not* invariant across ages: Older participants tended to receive lower total scores on the scale, independent of their level of systemizing, due to the presence of items relating to comfort with technology. This bias could be problematic if future studies wish to compare sum scores of groups with different mean ages, and may also be a problem for other longer systemizing scales. The development of systemizing measures that exhibit invariance across age groups may be a useful avenue for future research.

In confirmation of our hypotheses, the brief SQ had expected correlations with other constructs. As predicted, males had significantly higher brief SQ scores than females. After accounting for age differences, the Cohen's *d* effect size for this difference was 0.64, indicating a medium effect (Cohen, 1988). This effect size was similar to or slightly lower than the sex differences found in other studies that have used longer versions of the SQ (*d* = 0.60[1] in study 1 of Baron-Cohen et al., 2003; *d* = .96 and 1.01 in the two studies of Ling et al., 2009; *d* = 0.44 in study 1 and *d* = 0.96 in study 2 of Nettle, 2007; *d* = 0.96 in

[1] By our calculation.

Wakabayashi et al., 2006[1]). As predicted, the correlation between mental rotation and SQ was positive and $r = .24$ was similar in magnitude to $r = .25$ found in Ling et al. (2009). The relationship between systemizing and trait anxiety was negative, as predicted on the basis of previous research (Pingault et al., 2012). These findings support the construct validity of the brief SQ.

For the purposes of discriminant validation, we tested the relationship between systemizing and childhood-recalled agreeableness and extroversion. As expected, we found no statistically significant relationships between these variables and brief SQ score. Further supporting discriminant validity, our analyses showed that when entered into an exploratory factor analysis along with the other scales used in this study, the items of the brief SQ all loaded on the same factor with no items from other scales loading on this factor. These findings further support the construct validation of the brief SQ.

The Cronbach's alpha reliability estimate of .72 for the brief SQ is comparable to if a little lower than the alphas of .79-.91 reported in studies using the 40-item SQ (Baron-Cohen et al., 2003; Ling et al., 2009; Wakabayashi et al., 2006). The use of a 7-point rather than 4-point response scale may have helped to retain a reasonable level of reliability, but almost inevitably a shorter scale will result in some loss of reliability. According to Nunnally's (1978) classic guidelines, a reliability value of greater than 0.7 is adequate for the "early stages" (p. 245) of research, but not for more advanced basic research or for practical decision-making about individuals. As such, the brief SQ has adequate reliability for use in exploratory projects, or where systemizing is not the primary focus of the given project, whereas in other contexts a longer version of the SQ may still be more suitable.

Possibilities for Revisions

Most items we used in the brief version of the SQ performed well in our study. Item 5 "I can remember large amounts of information about a topic that interests me e.g. flags of the

world, airline logos" had the smallest factor loading and contributed slightly negatively to reliability. If this item continues to show poor psychometric properties in future studies, users of the scale may wish to revise, replace, or remove this item.

Limitations and Future Directions

The sample of data reported in this study was not originally collected to address the research questions being addressed here. This means that some of the characteristics of the sample and data are not optimal given the purposes of the current study. For example, the differing ages of the male and female samples was an idiosyncratic feature of the sample that somewhat complicated the analyses of invariance across gender, as discussed above. Furthermore, the study did not follow the typical validation procedure for shortened scales of comparing the performance of the shorter version of the scale with the longer one in the same sample. Future research could assess this, although we have good reason to expect that the brief version of the SQ would perform similarly to longer versions. Firstly, the brief version of the SQ had similar reliability to the longer 40-item version of the questionnaire; secondly, it had similar relationships with other constructs as other studies have found using longer versions of the questionnaire; and thirdly, other studies have found strong correlations between the 40-item SQ and shorter versions (r = .88 with a 14-item version, Samson & Huber, 2010; r = .91 with a 18-item version, Ling et al., 2009; and r = .95 with a 25-item version, Wakabayashi et al., 2006). While this study found that the brief version of the SQ generally had expected relationships with other constructs, future research could also assess whether this version of the scale also has expected relationships with the theoretically-related Empathy Quotient and Autism Quotient.

It is also worth noting that this was a scale validation study, and not a scale development study. We assessed the psychometric properties of a brief SQ that had already been reported in the literature, with only slight modifications. In future research, it might be

possible to develop a brief SQ with improved reliability and validity by selecting items from the full-length SQ based specifically on their psychometric properties.

Conclusion

This study showed that the brief version of the SQ had good reliability, factorial validity, and construct validity. The factor structure we found was in accordance with the intent of the scale (i.e. to measure a single systemizing factor). This was the first study to assess the psychometric properties of a brief version of the SQ questionnaire and the first to indicate factorial measurement invariance between genders, suggesting that the gender differences on which the underlying empathizing-systemizing theory is based are not the result of the SQ measuring a different construct in each gender. The brief version of the SQ may be particularly useful when researchers are interested in an assessment of systemizing that is less burdensome on participants than longer versions.

References

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences, 6*, 248–254. doi:10.1016/S1364-6613(02)01904-6

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with Asperger syndrome or high–functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*, 361–374. doi:10.1098/rstb.2002.1206

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*, 230–258. doi:10.1177/0049124192021002005

Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*. Retrieved from http://www.jstatsoft.org/v45/i03/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

Cook, C. M., & Saucier, D. M. (2010). Mental rotation, targeting ability and Baron-Cohen's empathizing–systemizing theory of sex differences. *Personality and Individual Differences, 49*, 712–716. doi:10.1016/j.paid.2010.06.010

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44*(11 Suppl 3), S78–S94. doi:10.1097/01.mlr.0000245454.12228.8f

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The International Personality Item Pool and the future of

public-domain personality measures. *Journal of Research in Personality*, *40*, 84–96. doi:10.1016/j.jrp.2005.08.007

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. doi:10.1080/10705519909540118

Ling, J., Burton, T. C., Salt, J. L., & Muncer, S. J. (2009). Psychometric analysis of the systemizing quotient (SQ) scale. *British Journal of Psychology*, *100*, 539–552. doi:10.1348/000712608X368261

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *4*, 73–79. doi:10.1027/1614-2241.4.2.73

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149. doi:10.1037/1082-989X.1.2.130

Manning, J. T., Baron-Cohen, S., Wheelwright, S., & Fink, B. (2010). Is digit ratio (2D:4D) related to systemizing and empathizing? Evidence from direct finger measurements reported in the BBC internet survey. *Personality and Individual Differences*, *48*, 767–771. doi:10.1016/j.paid.2010.01.030

Morsanyi, K., Primi, C., Handley, S. J., Chiesi, F., & Galli, S. (2012). Are systemizing and autistic traits related to talent and interest in mathematics and engineering? Testing some of the central claims of the empathizing–systemizing theory. *British Journal of Psychology*, *103*, 472–496. doi:10.1111/j.2044-8295.2011.02089.x

Nettle, D. (2007). Empathizing and systemizing: What are they, and what do they contribute to our understanding of psychological sex differences? *British Journal of Psychology, 98*, 237–255. doi:10.1348/000712606X117612

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Pingault, J.-B., Pouga, L., Grèzes, J., & Berthoz, S. (2012). Determination of emotional endophenotypes: A validation of the Affective Neuroscience Personality Scales and further perspectives. *Psychological Assessment, 24*, 375–385. doi:10.1037/a0025692

Pornprasertmanit, S., Miller, P., Schoemann, A., Rosseel, Y., Quick, C., Garnier-Villarreal, M., & Selig, J. (2013). semTools (Version 0.4–0). Retrieved from http://cran.r-project.org/web/packages/semTools/index.html

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*, 1–15. doi:10.1016/S0001-6918(99)00050-5

Rosseel, Y. (2013). lavaan: Latent variable analysis (Version 0.5-14). Retrieved from http://lavaan.ugent.be/

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Samson, A. C., & Huber, O. W. (2010). Short German versions of empathizing and systemizing self-assessment scales. *Swiss Journal of Psychology, 69*, 239–244. doi:10.1024/1421-0185/a000028

Taylor, J. A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal Psychology*, *48*, 285–90.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599–601. doi:10.2466/pms.1978.47.2.599

Veale, J. F., Clarke, D. E., & Lomax, T. C. (2010). Biological and psychosocial correlates of adult gender-variant identities: New findings. *Personality and Individual Differences, 49*, 252–257. doi:10.1016/j.paid.2010.03.045

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41,* 321–327. doi:10.1007/BF02293557

Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Goldenfeld, N., Delaney, J., Fine, D., … Weil, L. (2006). Development of short forms of the Empathy Quotient (EQ-Short) and the Systemizing Quotient (SQ-Short). *Personality and Individual Differences, 41,* 929–940. doi:10.1016/j.paid.2006.03.017

Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., … Wakabayashi, A. (2006). Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain Research, 1079,* 47–56. doi:10.1016/j.brainres.2006.01.012

The Psychometric Properties of a Brief Version of the Systemizing Quotient

Veale, JF

2017-07