General Intelligence (g), ACT Scores, and Theory of Mind: (ACT)g Predicts Limited Variance Among Theory of Mind Tests

Thomas R. Coyle\textsuperscript{a*}, Karrie E. Elpers\textsuperscript{a}, Miguel C. Gonzalez\textsuperscript{a}, Jacob Freeman\textsuperscript{b}, Jacopo A. Baggio\textsuperscript{c,d}

\textsuperscript{a} Department of Psychology, University of Texas at San Antonio, USA
\textsuperscript{b} Department of Sociology, Social Work, and Anthropology, Utah State University, USA
\textsuperscript{c} Department of Political Science, University of Central Florida, USA
\textsuperscript{d} Sustainable Coastal Systems Cluster, National Center for Integrated Coastal Research, University of Central Florida, USA

\textbf{A B S T R A C T}

This study is the first to examine relations between general intelligence (g), non-g factors, and theory of mind (ToM) using structural equation modeling with multiple indicators of g and ToM. g was based on the subtests of the ACT, a college admissions test that is strongly g loaded, and ToM was based on the Reading the Mind in the Eyes Test and the Short Story Test (SST). g correlated strongly with a latent ToM factor (β = .65) and moderately with the two ToM tests (β = .34), which correlated modestly with each other (β = .27). The modest correlation between the ToM tests indicates that g predicted a small amount of variance among the ToM tests (7%) and suggests that the ToM tests had little in common. In addition, non-g residuals of the ACT subtests, obtained after removing g, correlated negligibly with the ToM factor and the ToM tests (|β| < 0.06). Similar results were obtained for the ToM residuals, which correlated trivially with the ACT subtests. The trivial non-g effects suggest that g-ToM relations were attributable to “not much more than g.” The results replicated with different combinations of ACT subtests, controls for possible confounds (reading comprehension on the SST), and another college admissions test (the SAT). The use of a convenience sample (college students) and the limited measures of g and ToM are discussed as limitations. Future research should examine the robustness of effects using different measures of g and ToM and also examine possible mediators of g-ToM relations (e.g., executive functions).

1. Introduction

This study examines relations between general intelligence (g) and theory of mind (ToM), two constructs that appear conceptually distinct but are empirically related in adults (e.g., Baker, Peterson, Pulos, & Kirkland, 2014; see also, Botting & Conti-Ramsden, 2008; Estes & Bartsch, 2017; Fernández-Berrocal, Cabello, & Gutiérrez-Cobo, 2017; Henry et al., 2009). Whereas g reflects the ability to handle complexity and novelty (Gottfredson, 1997), ToM reflects the ability to infer the thoughts and feelings of other people (Baron-Cohen, 1995). Both g and ToM are needed to navigate complex interactions in diverse situations (e.g., Astington, 1998; Gottfredson, 1997) and both constructs are related to diverse cognitive abilities (e.g., executive functions) (e.g., Royall & Palmer, 2014; see also, Bull, Phillips, & Conway, 2008; Peterson & Miller, 2012).

The aim of the study was to characterize relations of ToM with g and non-g factors. g represents variance common to cognitive tests and largely explains the predictive power of tests (e.g., Jensen, 1998, pp. 270–305). Non-g factors represent factors unrelated to g and include non-g residuals of tests, obtained after removing g. Non-g factors measure specific abilities such as math or verbal abilities on scholastic aptitude tests (e.g., Coyle, Purcell, Snyder, & Kochunov, 2013; see also, Coyle & Pillow, 2008). Compared to g, non-g factors generally have weak predictive validity for work and school outcomes (e.g., Jensen, 1998, pp. 270–305).

A relationship between g and ToM would be predicted by Spearman’s principle of the indifference of the indicator (Jensen, 1998, pp. 32–34). The principle assumes that tests of all cognitive abilities are indicators of g (to varying degrees), regardless of the content or types of tests. The principle is bolstered by factor analysis of diverse tests, which typically correlate more strongly with g than with any other factor (e.g., Jensen, 1998, pp. 73–81). Spearman’s principle suggests that g should be linked to ToM because tests of ToM correlate with tests of executive functions and other diverse abilities, which are indicators of g (e.g., Royall & Palmer, 2014; see also, Bull et al., 2008; Peterson & Miller, 2012).

Spearman’s principle of the indifference of the indicator is related to what might be called the primacy of g hypothesis (cf. Jensen, 1984; see also, Reeve & Charles, 2008). The hypothesis assumes that relations among...
cognitive tests are primarily explained by g and that non-g factors (obtained after removing g) have limited effects. The primacy of g hypothesis extends Spearman's principle by identifying g as the primary source of variance that contributes to relations among tests. Support for the hypothesis comes from research showing that g explains most of the variance among diverse tests (e.g., verbal, math, spatial) and that non-g factors have relatively weak effects (e.g., Jensen, 1998, pp. 107–144). Based on the primacy of g hypothesis, non-g factors of tests were expected to correlate relatively weakly with ToM, which was expected to correlate strongly with g.

The relationship between g and ToM was examined using multiple tests. g was based on the subtests of the ACT (formerly, American College Test), a college admissions test widely used in the United States. The ACT correlates moderately with college grade point average (GPA) (e.g., Coyle et al., 2015; Coyle et al., 2013; Coyle, Purcell, Snyder, & Richmond, 2014; Coyle, Snyder, Richmond, & Little, 2015). In particular, math residuals predict achievements in science, technology, engineering, and math (STEM), while verbal residuals predict criteria in the humanities (e.g., English, art, music) (e.g., Coyle et al., 2015; see also, Coyle et al., 2013, 2014). Although non-g residuals typically have little or no predictive validity (Jensen, 1998, p. 30), the effect sizes of ACT non-g residuals for STEM and humanities criteria are substantial (β ≈ 0.30, Coyle et al., 2013, p. 118).

The non-g residuals of the ACT permitted a test of the specificity doctrine, which assumes that specific skills and abilities predict theoretically related criteria beyond g (cf. Jensen, 1984). The non-g residuals of the ACT subtests represent specific math and verbal abilities, obtained after removing g. The verbal residuals of the ACT were expected to predict ToM, which has been linked to diverse language and verbal abilities (e.g., vocabulary and reading comprehension) (e.g., Peterson & Miller, 2012; see also, Milligan, Astington, & Dack, 2007; Olderbak et al., 2015), suggesting that verbal abilities may be especially sensitive to ToM.

ToM was measured using the Reading the Mind in the Eyes Test (RMET; Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), and the Short Story Test (SST; Dodell-Feder, Lincoln, Coulson, & Hooker, 2013). The RMET is a social-perceptual test of ToM that presents participants with different pairs of eyes and asks them to identify the emotion being expressed (Baron-Cohen et al., 1997, 2001; Nettle & Liddle, 2008). In contrast, the SST is a social-cognitive test of ToM that presents participants with a short story and asks them to infer the thoughts and emotions of the characters (Dodd-Feder et al., 2013).

The two ToM tests are assumed to reflect different types of processing. The RMET is assumed to tap relatively automatic processing (Baron-Cohen et al., 2001), which is fast, intuitive, and effortless. In contrast, the SST is assumed to tap relatively effortful processing (Dodell-Feder et al., 2013), which is slower, reflective, and controlled. The SST and RMET correlate robustly with each other (r = 0.49, Dodell-Feder et al., 2013, p. 7), reflecting the fact that both tests tap ToM but measure different types of processing. In addition, both tests correlate positively with IQ and other verbal residuals were expected to correlate (relatively) strongly with ToM, whereas the SST residuals were expected to correlate (relatively) strongly with ToM, whereas the SST residuals were expected to correlate (relatively) strongly with ToM.

In the current study, the ToM measures (RMET and SST) were correlated with g based on the ACT subtests. All measures were drawn from a larger study of relations among general intelligence, social intelligence, and cooperation. The relations were estimated using structural equation modeling (SEM). By convention, g (variance common to tests) was estimated using all tests, while a latent ToM factor (loading on g) was estimated using the ToM tests (RMET and SST). SEM estimated relations between g, the ToM factor, and the ToM tests, which were also correlated with the non-g residuals (i.e., unique variances) of the ACT subtests.

Based on prior research linking measures of g with ToM (e.g., Baker et al., 2014; Dodell-Feder et al., 2013), g was expected to correlate positively with the ToM factor. In addition, g was expected to correlate more strongly with the SST, a social-cognitive measure of ToM, than with the RMET, a social-perceptual measure of ToM. The latter prediction was based on the assumption that the SST and g primarily measure cognitive processing, whereas the RMET primarily measures perceptual processing (cf. Nettle & Liddle, 2008). In addition, the non-g residuals of the ACT subtests, obtained after removing g, were expected to show a pattern of differential validity. Based on research linking ToM with verbal ability (e.g., Peterson & Miller, 2012), the verbal residuals were expected to correlate (relatively) strongly with ToM, whereas the math residuals were expected to correlate weakly (and nonsignificantly) with ToM. Such a pattern would be the first to demonstrate that g based on a widely used college admissions test (the ACT) predicts ToM, and that the verbal residuals of the test uniquely predict ToM.

2. Method

2.1. Subjects

Subjects were 374 college students who participated in a study of human intelligence, social intelligence, and cooperation.1 249 subjects had scores for three ACT subtests (math, reading, English), which were used in the primary analyses. A small number of subjects (29 of 374) had RMET and SST scores but no ACT scores and another set of subjects (96 of 374) had SAT scores for only two subtests (math and verbal). (The SAT is another college admissions test.)

---

1. The current study used only the measures of g (ACT subtests) and ToM (RMET, SST) from the original study, which also included measures of cooperation on a multiplayer video game. The measures of cooperation were not relevant for the current study and were excluded from the analyses.
test used in the United States.) Because three or more subtests were needed to identify the SEM model and estimate the relations for non-\(g\) residuals, the SAT data were not used in the primary analyses. To address the possibility that missing data distorted the estimates in the primary analyses (using only cases with ACT scores), supplemental analyses were performed using the full sample with all 374 subjects (see Statistical analyses).

2.2. Variables

The RMET and SST were presented on computers to small groups of subjects, with 8 to 20 subjects per group. ACT scores were obtained from university records. Descriptive statistics for all variables are reported in Table 1.

ACT scores (possible range = 0 to 36) were available for three subtests: math (ACTm), reading (ACTr), and English (ACTe) (\(N = 249\)). The reading and English subtests were used to measure verbal ability. ACT scores showed no signs of distributional anomalies (|skewness| < 0.12, |kurtosis| < 0.79) (Table 1).

RMET scores (possible range = 0 to 36) were available for all subjects (\(N = 374\)). The RMET was based on the task developed by Baron-Cohen et al. (2001). Subjects were presented with 36 pairs of eyes (one at a time) and were shown four verbal descriptions of the emotion conveyed by the eyes (e.g., happy, sad, angry, depressed). Subjects had to select the emotion conveyed by each pair of eyes.

SST scores (possible range = 0 to 16) were missing for eight subjects (\(N = 366\)). (Missing data were due to response omissions.) The SST was based on the task developed by Dodell-Feder et al. (2013). Subjects had to read a short story by Ernest Hemingway (The End of Something) and answer questions about the mental states (e.g., beliefs, intentions) of the characters in the story. In addition, subjects had to answer a series of questions about non-mental states in the stories, which were designed to control for reading comprehension. (The comprehension questions were used to check the robustness of results in the primary analyses.) SST and RMET scores showed no signs of distributional anomalies (|skewness| < 0.55, |kurtosis| < 0.35) (Table 1).

2.3. Statistical analyses

Statistical analyses were divided into primary and supplemental analyses. The primary analyses analyzed subjects with ACT scores (\(N = 249\)). Supplemental analyses analyzed all subjects (\(N = 374\)), including subjects with no ACT scores or ToM scores (RMET, SST), as well as a small sample of subjects with SAT scores (\(N = 96\)). Missing data were handled using full information maximum likelihood (FIML).

SEM with maximum likelihood estimation estimated relations between g and ToM. By convention, g was based on all tests and measured variance common to all tests. A ToM factor measured variance common to the ToM tests. \(g\) and ToM factor. Correlations were computed between \(g\) and the ToM factor, and between the non-\(g\) residuals of the ACT subtests (u1, u2, u3) and the ToM variables. The model depicts u1 \(\rightarrow\) RMET effect (\(\beta = -0.02\)), which represents the path coefficient between the ACTm residual and RMET (Table 4). (The base model omitted the u1 \(\rightarrow\) RMET path.)

Comparative fit index (CFI), root mean square error of approximation (RMSEA), and Akaike information criterion (AIC) (Kline, 2005, pp. 133–145). Model fit for the base model was excellent (\(\chi^2 = 0.35, df = 4, p = .99, CFI = 1.00, RMSEA = 0.00, AIC = 32.35\)) and was better than for a model in which the ToM tests loads directly on \(g\) with no ToM factor (\(\chi^2 = 7.75, df = 5, p = .17, CFI = 0.994, RMSEA = 0.047, AIC = 37.75\)).

The path from \(g\) to ToM (\(g\) \(\rightarrow\) ToM) estimated relations between \(g\) and ToM (Fig. 1). In addition, relations among the non-\(g\) residuals of the ACT math and verbal subtests, obtained after removing \(g\), were estimated by adding a path from a subtest’s residuals (u’s in Fig. 1) to a ToM variable (ToM factor, RMET, SST). Based on prior research (e.g., Peterson & Miller, 2012), the non-\(g\) effects were expected to be relatively strong (and significant) for the verbal residuals but weak (and nonsignificant) for the math residuals. All effects are reported as standardized coefficients, with the mean of multiple coefficients reported in parentheses (e.g., \(M_{r}\) and \(M_{M}\)). Following Cohen (1988, pp. 77–81), effects (or \(f^2\)) around 0.10, 0.30, and 0.50 were labeled weak, moderate, and strong, respectively. Significant effects are reported at \(p < .05\).

3. Results

3.1. Preliminary analyses

Table 2 reports bivariate correlations between the five manifest variables (ACTm, ACTr, ACTe, RMET, SST) using data from subjects with ACT scores (\(N = 249\)), which were also used in the primary analyses. The correlations showed positive manifold (i.e., positive correlations) (\(M_{r} = 0.39\), which is the basis of \(g\) (i.e., variance common to tests). The positive correlations indicate that people who performed well on one test generally performed well on all others. ACT scores correlated more strongly with each other (\(M_{r} = 0.67\) than with the ToM tests (\(M_{r} = 0.28\)). The ToM tests correlated moderately with each other (\(r = 0.26\)) and with the ACT subtests (\(M_{r} = 0.23, 0.28, 0.32, \text{math, reading, and English, respectively}\).

---

**Table 1**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTm</td>
<td>249</td>
<td>22.56</td>
<td>4.43</td>
<td>13.00</td>
<td>34.00</td>
<td>0.12</td>
</tr>
<tr>
<td>ACTr</td>
<td>249</td>
<td>24.83</td>
<td>5.66</td>
<td>12.00</td>
<td>36.00</td>
<td>-0.03</td>
</tr>
<tr>
<td>ACTe</td>
<td>249</td>
<td>24.02</td>
<td>5.41</td>
<td>10.00</td>
<td>36.00</td>
<td>0.06</td>
</tr>
<tr>
<td>RMET</td>
<td>249</td>
<td>27.55</td>
<td>3.51</td>
<td>18.00</td>
<td>34.00</td>
<td>-0.55</td>
</tr>
<tr>
<td>SST</td>
<td>245</td>
<td>9.29</td>
<td>2.61</td>
<td>1.00</td>
<td>15.00</td>
<td>-0.49</td>
</tr>
</tbody>
</table>

Note. Primary Analyses = subjects with ACT scores. Supplemental Analyses = all subjects. The descriptive statistics for the ACT scores were the same in both analyses and therefore are reported only in the Primary Analyses. Min = minimum observed score. Max = maximum observed score. ACTm = ACT math. ACTr = ACT reading. ACTe = ACT English. RMET = Reading the Mind in the Eyes Test. SST = Short Story Test.
relations were estimated using the SEM model (Fig. 1) with cases with ACT scores (N = 249). ACTm = ACT math, ACTr = ACT reading, ACTe = ACT English. RMET = Reading the Mind in the Eyes Test. SST = Short Stories Test.

### 3.2. Primary analyses

The analysis of effects of the ACT residuals after reversing the direction of effect between g and the ToM factor in the SEM model (Fig. 1). The new model was identical to the prior model except that the effect between g and the ToM factor was changed from g→ToM to g←ToM. As before, the analyses were performed separately for each ACT subtest (ACTm, ACTr, ACTe) and ToM variable (ToM factor, RMET, SST) and yielded a total of nine effects (3 ToM variables x 3 ACT subtests). Consistent with the prior analyses (cf. Table 4, Analyses 2 to 10), the ACT residuals (ACTm, ACTr, ACTe) correlated negligibly with the ToM factor (β = 0.05, 0.09, −0.10). In addition, all ToM variables correlated negligibly with the residuals of the ACT math subtest (Mβg = −0.02) and with the residuals of the ACT reading and English subtests (Mβg < −0.01).3

### Table 2

Correlations among manifest variables.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ACTm</td>
<td>−</td>
<td>0.56</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>2. ACTr</td>
<td>0.69</td>
<td>0.77</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>3. ACTe</td>
<td>0.24</td>
<td>0.30</td>
<td>0.34</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>4. SST</td>
<td>0.22</td>
<td>0.25</td>
<td>0.30</td>
<td>0.26</td>
<td>−</td>
</tr>
</tbody>
</table>

Note. All correlations p < .01. Correlations computed using only subjects with ACT scores (N = 249). ACTm = ACT math. ACTr = ACT reading, ACTe = ACT English. RMET = Reading the Mind in the Eyes Test. SST = Short Stories Test.

### Table 3

SEM implied correlations among g, ToM factor, ToM tests, and ACT subtests.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. g</td>
<td>−</td>
<td>0.65</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>2. ToM</td>
<td>0.35</td>
<td>0.55</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>3. RMET</td>
<td>0.32</td>
<td>0.49</td>
<td>0.27</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>4. SST</td>
<td>0.71</td>
<td>0.46</td>
<td>0.25</td>
<td>0.23</td>
<td>−</td>
</tr>
<tr>
<td>5. ACTm</td>
<td>0.80</td>
<td>0.51</td>
<td>0.28</td>
<td>0.25</td>
<td>0.57</td>
</tr>
<tr>
<td>6. ACTr</td>
<td>0.97</td>
<td>0.62</td>
<td>0.34</td>
<td>0.31</td>
<td>0.69</td>
</tr>
<tr>
<td>7. ACTe</td>
<td>0.97</td>
<td>0.62</td>
<td>0.34</td>
<td>0.31</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Note. All correlations p < .01. Correlations estimated using SEM and subjects with ACT scores (N = 249). g = latent factor based on all tests. ToM = latent factor based on RMET and SST. RMET = Reading the Mind in the Eyes Test. SST = Short Story Test. ACTm = ACT math. ACTr = ACT reading, ACTe = ACT English.

### Table 4

SEM relations between g, non-g residuals of ACT subtests, and ToM variables.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. g→ToM</td>
<td>0.65</td>
<td>0.35</td>
<td>(4, 0.99)</td>
<td>1.00</td>
<td>0.00 (1.00)</td>
</tr>
<tr>
<td>2. uACTm→ToM</td>
<td>−0.03</td>
<td>0.25</td>
<td>(3, 0.97)</td>
<td>1.00</td>
<td>0.00 (0.99)</td>
</tr>
<tr>
<td>3. uACTr→ToM</td>
<td>0.05</td>
<td>0.09</td>
<td>(3, 0.99)</td>
<td>1.00</td>
<td>0.00 (1.00)</td>
</tr>
<tr>
<td>4. uACTe→ToM</td>
<td>−0.06</td>
<td>0.30</td>
<td>(3, 0.96)</td>
<td>1.00</td>
<td>0.00 (0.98)</td>
</tr>
<tr>
<td>5. uACTm→RMET</td>
<td>−0.02</td>
<td>0.27</td>
<td>(3, 0.97)</td>
<td>1.00</td>
<td>0.00 (0.99)</td>
</tr>
<tr>
<td>6. uACTr→RMET</td>
<td>0.03</td>
<td>0.11</td>
<td>(3, 0.99)</td>
<td>1.00</td>
<td>0.00 (1.00)</td>
</tr>
<tr>
<td>7. uACTe→RMET</td>
<td>−0.03</td>
<td>0.30</td>
<td>(3, 0.96)</td>
<td>1.00</td>
<td>0.00 (0.98)</td>
</tr>
<tr>
<td>8. uACTm→SST</td>
<td>−0.01</td>
<td>0.34</td>
<td>(3, 0.95)</td>
<td>1.00</td>
<td>0.00 (0.98)</td>
</tr>
<tr>
<td>9. uACTr→SST</td>
<td>0.01</td>
<td>0.33</td>
<td>(3, 0.95)</td>
<td>1.00</td>
<td>0.00 (0.98)</td>
</tr>
<tr>
<td>10. uACTe→SST</td>
<td>−0.01</td>
<td>0.35</td>
<td>(3, 0.95)</td>
<td>1.00</td>
<td>0.00 (0.98)</td>
</tr>
</tbody>
</table>

Note. All non-g residual effects (u’s) are nonsignificant. Primary Analyses = subjects with ACT scores. Supplemental Analyses = all subjects. g = variance common to all tests. ToM = variance common to ToM tests. u = non-g residual (unique) variance of ACT subtests, obtained after removing g. ACTm = ACT math. ACTr = ACT reading, ACTe = ACT English. RMET = Reading the Mind in the Eyes Test. SST = Short Story Test.

Note. All non-g residual effects (u’s) are nonsignificant. Primary Analyses = subjects with ACT scores. Supplemental Analyses = all subjects. g = variance common to all tests. ToM = variance common to ToM tests. u = non-g residual (unique) variance of ACT subtests, obtained after removing g. ACTm = ACT math. ACTr = ACT reading, ACTe = ACT English. RMET = Reading the Mind in the Eyes Test. SST = Short Story Test.
3.3. Supplemental analyses

Table 4 (Analysis 11–20) reports supplemental analyses of the full sample (N = 374) using the same methods and the base model (Fig. 1). (These analyses included 125 subjects without ACT scores.) These analyses estimated the relations between g and the ToM factor (Analysis 11), and the relations between the non-g residuals of the ACT subtests and the ToM variables (ToM factor, RMET, SST) (Analysis 12–20). The results replicated the primary analyses (cf. Table 4, Analysis 1 to 10); g correlated strongly with the ToM factor (β = 0.64). In addition, the ToM factor and ToM tests correlated negligibly with the non-g residuals of all ACT subtests (Mg = −0.01) and with the residuals of the ACT math (Mg = −0.02) and verbal (Mg < −0.01) subtests. The mean difference in analogous effects for primary and supplemental analyses was trivial (M difference < |0.01|).

3.4. Robustness checks

Additional analyses controlled for SST comprehension of non-mental states (SSTC). These analyses were identical to the prior analyses (Table 4, Analyses 1–20), with one exception: An observed variable for SSTC was added to the SEM model (Fig. 1), which also included a path from the comprehension measure to SST reasoning (SST). The path from SSTC to SST controlled for story comprehension. The results replicated the prior analyses (Table 4, Analyses 1–20). g correlated strongly with the ToM factor (Mg = 0.63), and the ACT non-g residuals correlated negligibly with the ToM factor and ToM tests (|Mg| < 0.04). The mean difference in analogous effects with and without the SSTC in the model was trivial (M difference < 0.002).

Other analyses examined whether the relations between g and the ToM factor (g→ToM) were affected by the set of ACT subtests. The set included one math subtest and two verbal subtests (reading and English), one of which (English) had a very strong g loading (loading = 0.97, Table 3). To examine whether the verbally-loaded ACT subtests affected g→ToM relations, additional analyses estimated g using the ACT math subtest plus either the reading or English subtest. These analyses were identical to the primary analyses, except that only two ACT subtests loaded on g: the math subtest plus one verbal subtest (reading or English). (Like the primary analyses, these analyses were performed using only subjects with ACT scores.) Consistent with the primary analyses (cf. Table 3), g→ToM relations were strong when g was based on the math and reading subtests (β = 0.65) or when g was based on the math and English subtests (β = 0.63).7 The average difference in analogous effects for the current analyses (using two subtests) versus the original analyses (using three subtests) was trivial (M difference = −0.01) (cf. Table 3).

4. Discussion

This study was the first to examine g-ToM relations using a ToM factor based on multiple ToM tests (SST, RMET) and a g based on the ACT, a widely-used college admissions test. g correlated strongly with the ToM factor (β = 0.65) and moderately with the ToM tests (Mg = 0.34), which correlated modestly with each other (β = 0.27) (Table 3). The modest correlation between ToM tests indicates that g predicted limited variance among the ToM tests and suggests that the ToM tests had little in common. Moreover, the non-g residuals of the ACT math and verbal subtests, obtained after removing g, correlated negligibly with the ToM factor and ToM tests (Mg = −0.01), with trivial differences in effects for the math and verbal residuals (M difference = −0.02) (Table 4). Similar results were obtained for the ToM residuals, which correlated negligibly (and non-significantly) with the ACT subtests (|Mg| effects < 0.01) (Footnote 6). The results support the primacy of g hypothesis, which predicts that g largely explains the relations among cognitive abilities (cf. Jensen, 1984; see also, Reeve & Charles, 2008).8

The trivial non-g effects replicated with (a) all three ACT scores (math, reading, English), (b) ACT math scores and one verbal score (reading or English), and (c) SAT scores. In addition, the trivial effects replicated when the path between g and the ToM factor was reversed (Footnote 5) and when the ToM residuals predicted the ACT subtests (Footnote 6). Together, the results suggest that the trivial non-g effects were robust to variation in the statistical model and residual source (ACT residuals vs. ToM residuals).

It should be emphasized that the strong g→ToM relation (β = 0.65, Table 3) was based on a modest correlation between the two ToM tests (β = 0.27, Table 3), indicating that g correlated with a small amount of shared variance between the tests (7%). The modest correlation between the ToM tests indicates that the tests were largely independent and argues against a strong general ToM factor based on variance common to the tests. In addition, the ToM factor was based on two ToM tests (RMET, SST), which may not be representative of other ToM tests, potentially limiting the generalizability of the findings. Given the modest correlation between the ToM tests, the interpretation of the strong g→ToM relation must acknowledge the limited variance among the ToM tests and the possibility that different ToM tests may yield different results. The issue of the weak relation between the ToM tests will be revisited below in the section on Limitations.

The results extend research on g-ToM relations. First, whereas prior research estimated g-ToM relations using IQ tests (e.g., Baker et al., 2014; Dodell-Feder et al., 2013), the current study estimated g-ToM relations using the ACT, a college admissions test widely used in the United States. The ACT correlates strongly with g (e.g., corrected r = 0.77, Koenig et al., 2008) and with the non-verbal Raven’s Matrices (corrected r = 0.75, Koenig et al., 2008). In the current study, a g based on the ACT correlated strongly with the ToM factor (β = 0.65, Table 3). The strong g→ToM relation is consistent with the hypothesis that g and its proxies (the ACT) facilitate the ability to make complex inferences in everyday life (Gottfredson, 1997). Such inferences include the ability to infer the mental states of others, which is a key component of ToM (Baron-Cohen, 1995).

Second, whereas prior research estimated g-ToM relations using the RMET (Baker et al., 2014) or SST (Dodell-Feder et al., 2013), the current study estimated g-ToM relations using a latent ToM factor based on the RMET and SST. The ToM factor measured variance common to both the RMET and SST (after removing variance unique to the tests). The correlation between the ToM factor and g was strong (β = 0.65, Table 3) and was much stronger than a meta-analytic correlation between IQ and the RMET (rIQ,RMET ≈ 0.24, Baker et al., 2014). The strong correlation in the current study can be attributed to the use of multiple ToM tests and the removal of unique variance from the individual ToM tests. The use of multiple tests provides a more valid estimate of g-ToM relations compared to individual ToM tests (RMET or SST), which are loaded with more unique variance and less common

---

7 Supplemental analyses using similar methods examined the independent group of subjects with SAT scores (N = 96), who received the same ToM tests. (The results should be interpreted with caution, given the relatively small number of subjects with SAT scores compared to number with ACT scores (N = 249).) These subjects had two SAT scores (math and verbal). The two scores were used to estimate relations between g and the ToM factor but not relations with the SAT non-g residuals, which would require three or more SAT scores to identify the SEM model and estimate effects of the non-g residuals. Consistent with analyses involving two ACT scores (i.e., ACT math plus reading or English), the g based on the two SAT scores correlated strongly with the ToM factor (β = 0.52), a large effect according to Cohen’s (1998) criteria.

8 It is worth noting that the average ratio of unique variance (based on the ACT non-g residual effects with ToM variables) over common variance (based on the g-ToM effect) was very small (0.04), further highlighting the strong effects of g and the weak effects of the non-g residuals (Table 4, Primary analyses).

It is worth noting that the RMET and SST measure different aspects of ToM. The RMET measures social-perceptual ToM (i.e., inferring emotions from eyes), which is assumed to be fast and automatic (Baron-Cohen et al., 1997, 2001). In contrast, the SST measures social-cognitive ToM (e.g., inferring mental states of story characters), which is assumed to be slow and effortful (Dodell-Feder et al., 2013). Because of its cognitive orientation, the SST was expected to correlate more strongly with the RMET than with g, a measure of cognitive ability. This hypothesis was not supported. The relations of g with each test (SST or RMET) were similar and not appreciably different (difference = 0.03, Table 3). The trivial difference in the relations of g with the SST and RMET suggests that, whatever their conceptual differences, both tests have similar loadings on g.

Third, whereas prior research estimated g-ToM relations (e.g., Baker et al., 2014; Dodell-Feder et al., 2013), the current study also estimated ToM relations with the non-g residuals of the ACT math and verbal subtests, obtained after removing g (cf. Coyle et al., 2013, 2015). Because ToM has been found to be especially sensitive to verbal abilities (e.g., Peterson & Miller, 2012), the verbal residuals were expected to predict ToM better than the math residuals. This hypothesis was not supported. The ACT verbal and math residuals showed negligible correlations with all ToM variables (|β| < 0.06) (Table 4), suggesting that specific cognitive abilities, obtained after removing g, contribute little to g-ToM relations. The negligible non-g effects, coupled with the strong g-ToM relations, indicates that ACT is linked to ToM via g (i.e., variance common to all tests). Stated differently, the link between ACT and the ToM is attributable to “not much more than g” (cf. Stauffer, Ree, & Carretta, 1996).

4.1. Limitations and future research

The current study has limitations that can guide future research. First, as noted above, the strong g-ToM relation (β = 0.65, Table 3) is based on a modest correlation between the two ToM tests (β = 0.27, Table 3), which indicates that g correlates with limited variance among the ToM tests (7%).9 The relatively small amount of variance shared by the two ToM tests suggests that the tests are empirically distinct and measure different aspects of ToM (e.g., cognitive and perceptual). Moreover, the modest relation among the ToM tests may reflect a weak relation among ToM tests in general. ToM tests measure diverse ToM components (e.g., cognitive or affective; verbal or visual) (Henry et al., 2013, pp. 829–830), which may share limited variance and be empirically distinct. To test this possibility, future research should analyze a broader set of ToM tests, which would yield a better estimate of ToM variance and increase the validity of the g-ToM effects.

A second limitation concerns the use of the ACT as a measure of g and specific abilities. The ACT correlates strongly with g (e.g., r = 0.77, Koenig et al., 2008) and with specific math and verbal abilities, which predict school and work criteria (e.g., Coyle et al., 2014). However, the ACT excludes non-academic and other abilities (e.g., spatial abilities), which may also predict ToM. Future research should examine models of intelligence that sample a broader range of abilities. Two such models are the Cattell-Horn-Carroll (CHC) model (McGrew, 2009) and the verbal-perceptual-rotational (VPR) model (Johnson & Bouchard, 2007). The VPR and CHC models incorporate spatial abilities, which may predict ToM (cf. Cook & Saucier, 2010), and emotional intelligence (e.g., MacCann, Joseph, Newman, & Roberts, 2014), which may also predict ToM (e.g., Ferguson & Austin, 2010).

A third limitation concerns the sample of college students, who were admitted to college after taking the ACT, a selection test used to screen out low ability applicants. Compared to non-admitted applicants, admitted applicants would be expected to have higher levels of cognitive ability and possibly ToM, which correlates with g. In the current study, higher levels of cognitive ability and ToM could restrict variance on g and ToM, which could attenuate g-ToM effects. Although the strong g-ToM effects replicated after correcting for range restriction (Footnote 3), future research should use samples with a wider range of ability and age, which predicts cognitive functions related to ToM and g (e.g., executive functions) (e.g., Bull et al., 2008). Such samples would increase variance on g and ToM and could be used to examine whether g-ToM effects are mediated by other factors (e.g., executive functions).

A final limitation, related to the use of higher ability subjects, concerns Spearman’s Law of Diminishing Returns (SLODR) (e.g., Jensen, 1998, pp. 585–588). SLODR is based on Spearman’s (1932) observation that the correlations among mental tests decrease at higher ability levels, a pattern confirmed by a recent meta-analysis (Blum & Holling, 2017). The decrease (in correlations) is related to decreases in the g loadings of tests and the predictive validity of tests (Jensen, 1998, pp. 274–294). Based on SLODR, the use of higher ability subjects would be expected to depress g-ToM relations, which should decline at higher ability levels. Future research should estimate g-ToM relations at different ability levels and examine whether g-ToM relations decline at higher ability levels, which is predicted by SLODR.

5. Conclusion

This research examined g-ToM relations using an SEM model based on multiple measures of g and ToM (ACT, SST, RMET). g was based on the subtests of the ACT, which is strongly g loaded, and ToM was based on the SST and RMET, which measure different aspects of ToM (viz., perceptual and cognitive). The results indicated that g correlated strongly with ToM (β = 0.65, Table 4), and that the non-g residuals of the ACT subtests correlated trivially with ToM (|β| < 0.06, Table 4). However, the strong g-ToM relation was based on a modest correlation between the ToM tests (β = 0.27, Table 3), indicating that g correlated with limited variance among the ToM tests (7%). Future research should examine the robustness of effects using different measures of g and ToM and also examine possible mediators of g-ToM relations (e.g., executive functions).

Author note

This research was supported by a grant from the National Science Foundation’s Interdisciplinary Behavioral and Social Science Research Competition (IBSS-L 1620457). The authors thank three anonymous reviewers for their constructive comments.

References


