Gs Invaders:
Assessing a computer game-like
test of processing speed

Jason McPherson and Nicholas R. Burns
University of Adelaide, Adelaide, South Australia, Australia

Computer games potentially offer a useful research tool for psychology but there has been little use made of them in assessing cognitive abilities. Two studies assessing the viability of a computer game-like test of cognitive processing speed are described. In Experiment 1, a computerized coding task that uses a mouse response method (McPherson & Burns, 2005) was the basis for a simple computer game-like test. In Experiment 2, dynamic game-like elements were added. Validity was assessed within a factor analytic framework using standardized abilities tests as marker tests. We conclude that computer game-like tests of processing speed may provide an alternative or supplementary tool for research and assessment. There is clearly potential to develop game-like tests for other cognitive abilities.

Psychologists have used computers for at least three decades, with uses ranging from the automation of simple questionnaire items to assistance in taking complex psychological histories (Epstein & Klinkenberg, 2001). Computerization of clinical instruments, personality scales, job attitude scales and various types of cognitive tests has lead to many advantages including ease of administration and standardization, less training for proctors, fewer scoring errors, faster (even immediate) results, fewer disposable materials and, in some cases, fewer opportunities for cheating (Epstein & Klinkenberg, 2001; Mead & Drasgow, 1994). In the area of cognitive abilities assessment, the potential of these advantages has been a motivating force behind large-scale projects such as the computerized batteries developed by the U.S. military (Alderton, Wolfe, & Larson, 1997; Peterson et al., 1990; Ree & Carretta, 1998) and by various educational institutions (Bennett, 1998; Mead & Drasgow, 1994). Furthermore, many potential developments in abilities testing have been identified that are largely dependent on tests being computerized (Embretson, 2003). It would thus seem that computerization is certain to play a major role in abilities testing in the 21st century.

Most computerized abilities testing has involved the translation of traditional paper-and-pencil (PP) tests to computer format and it appears that adequately designed computerized tests are potentially as valid as traditional PP tests (Embretson, 2003; Epstein & Klinkenberg, 2001; Mead & Drasgow, 1994; Van de Vijver & Harsveld, 1994). However, given that modern computers are capable of much more dynamic presentation of stimuli than is the case with PP tests, it seems a limited utilization of a very powerful tool simply to imitate via computer the administration of tests not designed with computerization in mind. Better use of the distinctive qualities of computer technology may bring additional advantages for both test administrators and testees alike.

One format that does make use of the distinctive features of computerization is the computer game (Donchin, 1995; Porter, 1995; Washburn, 2003). Although some researchers interested in studying cognitive processes have used computer game formats (Berger, Jones, Rothbart, & Posner, 2000; Donchin, 1995; Graham, Cook, Cohen, Phelps, & Gerkovich, 1985; Porter, 1995; Ryan, 1994), there appears to have been less research investigating their use in assessing individual differences in cognitive abilities. Rabbitt, Banerji, and Szymanski (1989) studied the relationship between IQ and various measures obtained from a computer game called Space Fortress. Space Fortress was developed to assess the role of practice in skill acquisition (Donchin, 1995) and involves a rather complex set of operations. Rabbitt et al. found that performance on Space Fortress was related to IQ, but only after a substantial amount of practice on the game. They argued that practice was necessary to overcome any preexisting differences in familiarity with computer games. Jones, Dunlap, and Bilodeau (1986) assessed the relationship between five arcade style video games and a battery of abilities tests and found that game performance was moderately related to all of the abilities tests they used except verbal ability tests. Similar results were obtained by Jackson, Vernon, and Jackson (1993) who found that a customized computer game requiring dynamic spatial tracking loaded on a factor with the Performance, but not the Verbal, tests from the Multidimensional Aptitude Battery.

J. McPherson, jlmcpher@psychology.adelaide.edu.au

Although these studies suggest a relationship between performance on computer games and cognitive abilities, we believed it would prove more fruitful to design a computer game-like test from first principles with the intention of measuring specific cognitive abilities. Such systematic use of computer games to measure psychometric constructs seems to be limited to researchers who have used game-like elements in traditional tests such as Anderson's (1992) use of alien characters when measuring inspection time in children. It is likely that many other researchers have employed game-like elements such as this to engage the interest of children (Berger et al., 2000), but none of these appear to have been further developed or validated as psychometric measures.

The goal of the present study was to assess the construct validity of a simple computerized game-like test. Although the distinction between computer based tasks and computerized games is probably a blurred one (Washburn, 2003), we used game-like graphics and sound, a cumulative scoring and feedback system, and a relevant cover story to provide a more game-like experience for users. We based this game-like test on our simple computerized coding task (McPherson & Burns, 2005) tentatively identified as a measure of Processing Speed (Gs) within the framework of the Cattell–Horn–Carroll (CHC) theory of cognitive abilities (McGrew & Flanagan, 1998). We believe CHC provides an ideal framework for evaluating new tests because many traditional tests have been identified as marker tests for the broad ability domains described within this theory.

CHC theory provides a structural model of human cognitive abilities representing the main points of convergence within modern abilities research (McGrew & Flanagan, 1998). The model incorporates at least two strata but arguably three. The first stratum comprises over 60 abilities that are narrow and which correspond conceptually to Thurstone's Primary Mental Abilities (1938; 1947). The pattern of intercorrelations among these first-stratum abilities defines about ten broad abilities at the second stratum. These second-stratum abilities include fluid and crystallized intelligence, as first described by Cattell (1943; 1963) along with others more recently described, primarily by Horn and Noll (1997). Controversy remains on the interpretation of the pattern of intercorrelations among these broad second-stratum abilities. Many researchers and theorists argue for the existence and primacy of a third-stratum general factor (g; see Carroll, 1993; Jensen, 1998) while others dismiss the importance of g (e.g., Stankov, 1998). Deary and Caryl (1997) make the point that whether the broad second-stratum abilities or the putative third-stratum general factor are considered most important depends on the particular research question at hand. Considering individual tests are generally taken to represent the first-stratum the goals of the present study seemed best suited to analysis at the second-stratum.

In summary, we aimed to assess whether the game-like coding task would measure Gs. Considering our use of fairly complex graphics as stimuli, it also seemed appropriate to assess how game performance might also relate to the domain of Visual Processing (Gv). These broad aims were addressed in two separate experiments. In Experiment 1, our simple computerized coding task (McPherson & Burns, 2005) was altered by substituting more game-like graphics and sound. In Experiment 2, this modified test was expanded to incorporate some of the more dynamic elements common to computer games. Experiment 2 also incorporated single marker tests for two other broad abilities to assess discriminant validity. The rationale for conducting two separate experiments was to first assess the effect of the more complex game-like visual stimuli before adding more dynamic game elements.

EXPERIMENT 1

Method

Participants

Full data sets were obtained for $N = 60$ participants. Thirty-five participants were Level I Psychology students from the University of Adelaide who received course credit for their participation, and 25 were students recruited more generally from the university and paid $20(AU) for their participation. The mean age for the whole sample was 20.6 years ($SD = 2.81$); 32 were female (mean age = 20.5 years, $SD = 2.27$), and 28 were male (mean age = 20.8 years, $SD = 3.36$).

Materials
Tests of Gs
Digit Symbol. This test from the Wechsler Adult Intelligence Scale (WAIS–III) requires the filling in of blank cells according to a key provided at the top of each test sheet. The standard 2-min time limit was used.

Visual Matching. This test, taken from the Woodcock–Johnson III Tests of Cognitive Abilities (WJ–III COG; Woodcock, McGrew, & Mather, 2001) requires searching for and circling matching pairs of numbers. The standard 3-min time limit was used.

Tests of Gv
Block Design. This test from the WAIS–III requires the replication of two-colored models using blocks. The designs are progressively difficult, and scores reflect both accuracy and speed.

Spatial Relations. This test is taken from the WJ–III COG (Woodcock, McGrew, & Mather, 2001). The participant is required to identify the two or three pieces that make up a larger target shape.

Custom Computer Tests
Symbol Digit. The layout for this computerized test is as described in detail elsewhere (McPherson & Burns, 2005). Each item consists of a symbol presented in the center of the screen situated within a white box with a black border (5.4×5.2 cm). Participants respond by left clicking the mouse on a 3×3 numerical grid (6.8×6.2 cm) in the bottom right hand corner of the screen. To make the use of this number grid easier, each number's background turns red when the cursor is placed over it and the cursor is restricted such that it cannot be moved outside of the grid area. The correct response is determined according to the code table (22.1×4.7 cm) at the bottom left-hand side of the screen. When a correct response is made, a "laser" style sound is heard and the next item appears in the center of the screen. When an incorrect response is made the computer's default beep sound is heard and the item remains until the correct response is made. The background color of the screen was blue. Seven practice items were provided before the test began. Participants were instructed to complete as many items as possible in 2 min. The sequence of responses required is the same as for Digit Symbol.

Space Code. The screen layout for this test consisted of a spaceship cockpit with a numerical response grid of the same dimensions as that used for Symbol Digit and situated in the center of the screen. However, the number grid is dark blue with white digits. To make the use of this number grid easier, each number's background turns light blue when the cursor is placed over it and, as with Symbol Digit, the

cursor is restricted such that it cannot be moved outside of the grid area. Each item consists of a single spaceship (approximately 5 × 5 cm) appearing in the window view of the cockpit. At the bottom of the cockpit, the nine spaceships are each presented with a single digit placed directly above. Participants are instructed by onscreen text that to destroy a ship requires the firing of the number placed above the matching ship at the bottom of the screen. When a correct response is made a laser sound is heard followed by an audible and visible explosion of the spaceship. When an incorrect response is made the screen flashes and a "banging" sound indicates that the participant's ship has been hit. Seven practice items are provided before the test begins. Participants are instructed to be as fast and accurate as they can and to destroy as many spaceships as possible in 2 min. The sequence of responses required is the same as for Digit Symbol.

Mouse–Grid Speed. This test used the same response grid layout as used in Space Code (see above). Participants practiced clicking two simple patterned sequences of digits (digits 1–9 in order and 1-3-9-7-5) as quickly as they could for 20 sec. Following this, they were asked to remember a final sequence (2-6-8-4-5) before being asked to click this pattern as quickly as possible for 20 sec. The number of correct clicks in this final trial served as the dependent measure for this test. This test was included to provide a simple measure of how quickly participants could use the response grid independently of the cognitive demands of the coding operation.

All computerized tests were run on a Pentium 4 class computer (2.4 MHz) with a Microsoft optical Intellimouse, which was brand new at the start of the experiment. Display was via a 19-in. monitor with a vertical refresh rate of 100 Hz and a screen resolution of 1,280 × 1,024 pixels.

Procedure

Participants were tested individually in a quiet room and asked to concentrate as best they could for each individual test. It was further emphasized that it was important for them to do their best on each test so that an accurate comparison could be made across their performances on all tasks. Tests were administered to all participants in the following order: Visual Matching; Digit Symbol; Mouse–Grid Speed; Space Code; Block Design; Spatial Relations; and Symbol Digit. Sessions took approximately 1 h to complete.

Results

Descriptive Statistics

Descriptive statistics for all tests are presented in Table 1.

Correlations

Intercorrelations between all tests are presented in Table 2. As expected, the two Gs marker tests, Digit Symbol and Visual Matching, were moderately correlated, and the two Gv marker tests, Block Design and Spatial Relations, were also moderately correlated.

Table 1
Descriptive Statistics for Ability Test Scores and
Computer-Based Mouse Response Test Scores

Measure	Response Time (in Milliseconds)			
	M	*SD*	Minimum	Maximum
Digit symbol	89.8	11.8	67	125
Visual matching	54.7	6.3	44	69
Block design	54.7	8.3	34	68
Spatial relations	74.6	3.9	63	80
Symbol digit	84.9	14.7	63	132
Space code	62.6	9.2	43	83
Mouse–grid speed	56.7	11.2	39	92

Considering the computer based mouse response (CBMR) tests it can be seen that Symbol Digit and Space Code share a considerable amount of variance. It can also be seen that Mouse–Grid Speed, which was designed to measure mouse response speed independently of coding speed, shared a similar amount of variance with each of the two CBMR tests. However, the partial correlation between Symbol Digit and Space Code controlling for Mouse–Grid Speed was still .60 ($p < .001$), suggesting that mouse speed was not the primary determinant of performance in these tasks.

Exploratory Factor Analysis (EFA)

The construct validity of Symbol Digit and Space Code was further assessed using EFA. In addition to Symbol Digit and Space Code, the two Gs markers and two Gv markers were submitted to Principal Axis Factoring (PAF). PAF was used because the primary goal was to identify common factors. Mouse–Grid Speed was not included so as to reduce the possibility of the CBMR tests altering the measurement space too strongly by introducing extra method variance. The number of factors to extract was decided by inspection of the scree plot (Cattell, 1966), parallel roots analysis (Montanelli & Humphreys, 1976), and the theoretical issues relevant to the present study.

Inspection of the scree plot and the parallel roots analysis both indicated a two-factor solution. A two-factor solution was also deemed appropriate considering the theoretical expectations of separate Gs and Gv factors. The two-factor PAF solution was submitted to Promax rotation (Kappa = 4) which resulted in two weakly correlated factors ($r = .27$) each with at least three loadings larger than .5. This solution accounted for 57.7% of the variance. The structure matrix of factor loadings is presented in Table 3. The Pattern Matrix indicated almost identical results.

Using the marker tests as a primary guide it can be seen that the first factor is strongly loaded by both Gs marker tests and thus supports the interpretation of this factor as representative of this construct. Similarly, the second factor is strongly loaded by the two Gv tests and is correspondingly interpreted as a Gv factor. Looking at the loadings for the CBMR tests, Symbol Digit has a salient loading only on the Gs factor, whereas Space Code has substantial loadings on both.

Correlations With Composite Measures

To further assess the relationship between the CBMR tests and the underlying ability constructs, composite Gs and Gv scores were derived using only the traditional marker tests. Gs composites were derived by computing average z scores for Digit Symbol and Visual Matching. Gv composites were derived computing average z scores for Block Design and Spatial Relations. These results closely match those obtained through EFA. Mouse–Grid Speed was moderately correlated with Gs scores ($r = .33$, $p = .009$), but not with Gv scores ($r = .14$, $p = .289$). Similarly, Symbol Digit was more strongly related to Gs scores ($r = .55$, $p < .001$) than to Gv scores ($r = .22$, $p = .085$). However, Space Code was related to both Gs ($r = .55$, $p < .001$) and Gv scores ($r = .50$, $p < .001$).

Table 2
Correlation Matrix for Ability Test Scores and
Computer-Based Mouse Response Test Scores

Variable	1	2	3	4	5	6	7
1. Digit symbol	–						
2. Visual matching	.49**	–					
3. Block design	.16	.45**	–				
4. Spatial relations	−.10	.17	.55**	–			
5. Symbol digit	.53**	.41**	.25	.15	–		
6. Space code	.40**	.55**	.53**	.36**	.65**	–	
7. Mouse–grid speed	.19	.38**	.22	.21	.31*	.36**	–

$^* p < .05.$ $^{**} p < .01.$

It is notable that Symbol Digit and Space Code showed correlations with the Gs scores comparable to the correlation between the two traditional measures comprising the composite ($r = .49, p < .001$). Space Code also showed a correlation with Gv scores comparable to the correlation between the two Gv markers ($r = .55, p < .001$).

Discussion

Results from Experiment 1 support previous research indicating that the Symbol Digit test using a computer-based mouse response method can be interpreted as a measure of Gs (McPherson & Burns, 2005). However, although the Space Code test loaded on the same Gs factor it also had a loading of a similar magnitude on the factor loaded by the Gv marker tests. This raises the possibility that this test may be a mixed measure of Gs and Gv. However, considering that the demands of the task seem qualitatively different from the Gv marker tasks, and given the small sample size of the present study, we were cautious about this interpretation.

With this in mind, we decided to proceed with a second experiment incorporating as many dynamic game-like elements as possible into Space Code. We also increased the number of marker tests used for each construct, to provide a better sampling of the relevant domains, and assessed the test–retest reliability of the new game-like version. We also included a test requiring visual associative memory and a test of more general visual reasoning ability to address the possibility that visual memory and reasoning processes could be related to performance of a highly visual game-like task. Additionally, we endeavored to address the issue of a certain responding strategy that was noted as being potentially problematic in the first experiment. For the Space Code test some participants

Table 3
Structure Matrix for Principal Axis Factoring of
Ability Test Scores and Computer-Based Mouse Response
Test Scores With Promax Rotation (Kappa = 4)

Variable	Processing Speed (Gs)	Visual–Spatial (Gv)
Digit symbol	**.724**	.061
Visual matching	**.668**	.415
Block design	.433	**.769**
Spatial relations	.148	**.725**
Symbol digit	**.724**	.316
Space code	**.772**	**.637**

Note—Bold numbers indicate loadings > .60.   Factors are correlated .27 and named on the basis of marker tests.

were observed to respond very rapidly without concern for accuracy within the Space Code test. It was therefore deemed important to incorporate into the Space Code protocol some form of disincentive toward this strategy of responding.

EXPERIMENT 2

Method

Participants

Full data sets were obtained for $N = 61$ participants. All participants were Level I Psychology students from the University of Adelaide who received course credit for their participation. The mean age for the whole sample was 20.0 years ($SD = 3.56$); 37 were female (mean age = 20.3 years, $SD = 4.25$), and 24 were male (mean age = 19.7 years, $SD = 2.14$).

Materials
  Tests of Gs
  Digit Symbol. Described in Experiment 1.
  Visual Matching. Described in Experiment 1.
  Decision Speed. This test from the WJ–III COG requires searching for pairs of conceptually matched objects and was administered according to standard instructions. Together with Visual Matching, this test is used to derive the WJ–III COG Processing Speed cluster score.
  Cross Out. This test is taken from the Woodcock–Johnson Psycho-Educational Battery–Revised (Woodcock & Johnson, 1989) and requires testees to cross out five matching patterns in rows of distracting patterns. The standard 3-min time limit was applied.
  Tests of Gv
  Block Design. Described in Experiment 1.
  Spatial Relations. Described in Experiment 1.
  Picture Recognition. This test from the WJ–III COG requires the recognition of a subset of previously presented pictures within a field of distracting pictures.
  Additional abilities tests
  Raven's Advanced Progressive Matrices–Short Form (APM–SF). This test is a brief, 12-item version of the Raven's Advanced Progressive Matrices (Bors & Stokes, 1998), which is generally regarded as a measure of fluid reasoning ability (Gf). Items require the identification of a missing part required to complete a pattern or series of visual stimuli arranged in a matrix. A 15-min maximum time limit was allowed.
  Visual–Auditory Learning (VAL). From the WJ–III COG, this test requires testees to learn a pictographic language. Visual stimuli are presented as representing common words and testees are asked to read sentences constructed from these visual stimuli. The number of associations that must be stored in memory to complete sentences is increased throughout the test. This test is regarded as a measure of long-term storage and retrieval (Glr).
  Customized Computer Tests
  Symbol Digit. Described in Experiment 1.
  Space Code Game. This test has the same layout and dimensions as Space Code (described in Experiment 1) with the addition of vari-

ous computer game-like elements. (1) Space ships to be destroyed were programmed to move according to a random movement algorithm; however, no aiming was required. (2) A timer was situated in the top right-hand corner of the screen. The timer began at 120 sec and proceeded downward to indicate how much time was left. Additionally, a verbal warning was provided at 90, 60, 30, 10, and 5 sec. This was by way of male-pitched, synthesized speech provided under Windows XP (Microsoft Sam). (3) A scoring system was implemented and a cumulative total was presented in the top left of the screen. Bonus stars were also awarded at predetermined score levels, and these were presented above and below the score. Each star indicated that a larger number of points would be awarded for each ship destroyed. The number of points awarded for each ship was displayed just to the left of the mouse response grid. All of these elements were described in an introduction screen prior to commencement of the game. (4) In addition to the laser and explosion sounds accompanying a correct response, the score was incremented in steps and associated with a high-pitched, rising sound. (5) Incorrect responses were accompanied by an equivalent score decrement accompanied by a lower pitched, descending tone. Additionally, to make incorrect responses more aversive, the response grid was also deactivated during the time it took for the score to decrease. This meant that the participant would have to wait a short period before being able to fire again. It was hoped that this would discourage the rapid firing strategies used by some participants in Experiment 1. (6) Background music was played during the game to further enhance the game-like feel of the task. The dependent variable was the total number of correct responses.

Guided Mouse–Grid Speed. This test used the same response grid layout as that used in Mouse–Grid Speed (described in Experiment 1). However, to improve on the task used in the first study, this test provided a more extensive demonstration of the task, and patterns were additionally indicated by the response buttons' lighting up. Participants were told that their task was to click on patterns as fast as they could for 15 sec and that there would be an untimed practice run preceding each of four patterns, in which they could learn the pattern. They were also instructed that the next number in each sequence would additionally be indicated by the corresponding button lighting up (this was intended to remove the requirement of holding the pattern sequence in memory). The experimenter then performed a demonstration trial with the participant watching. The four numerical sequences completed by participants were [1,2,3,4,5,6,7,8,9], [1,9,7,3,5], [2,6,8,4,5], and [5,1,5,2,5,3,5,6,5,9,5,8,5,7,5,4]. If the wrong button was clicked, the current number button simply remained lit up. The dependent variable was the total number of correct clicks across all four patterns.

All computerized tests were run on a Pentium 4 class computer (2.4 MHz) with a Microsoft optical Intellimouse, which was brand new at the start of Experiment 1. Display was via a 19-in. monitor with a vertical refresh rate of 100 Hz and a screen resolution of 1,280 × 1,024 pixels.

Procedure

Participants were tested individually in a quiet room and were asked to concentrate as well as they could for each individual test. It was further emphasized that it was important for them to do their best on each test so that an accurate comparison could be made across their performances on all tasks. Test were administered to all participants in the following order: Guided Mouse–Grid Speed; Space Code Game; Digit Symbol; Visual Matching; Cross Out; Decision Speed; Symbol Digit; APM–SF; Visual–Auditory Learning; Block Design; Spatial Relations; Picture Recognition; and Space Code Game. The second administration of Space Code Game was to provide an estimate of test–retest reliability. Sessions took approximately 2 h to complete.

Results

Descriptive Statistics

Descriptive statistics for all tests are presented in Table 4. It can be seen that there was a substantial practice

effect across the two administrations of the Space Code game; a paired samples t test indicated that this difference was highly significant [$t(60) = 12.0, p < .001$].

Correlations

The test–retest correlation for Space Code was .84, indicating acceptable reliability for this measure. Intercorrelations for Gs and Gv marker tests and CBMR tests are presented in Table 5. As expected, the PP format Gs marker tests were all moderately to strongly correlated among themselves, but not as strongly with the Gv tests. However, looking at the relationships among the three Gv marker tests, it can be seen that, although Block Design and Spatial Relations share considerable variance with each other, Picture Recognition is not significantly correlated with either of these tests. This suggests that Picture Recognition was not a good indicator of Gv within the present sample. Looking at the CBMR tests, it can be seen that there is a positive manifold with all of the other ability tests—and particularly with the Gs marker tests. This supports the interpretation, that these tasks share common ability requirements with PP-format Gs measures.

Exploratory Factor Analysis

To further assess the construct validity of both the Symbol Digit and Space Code Game tests, EFA procedures were carried out with these two tests and the Gs and Gv marker tests. In order to balance the number of measures for each construct, Visual Matching and Decision Speed were used to derive the single Gs composite measure as indicated for the WJ–III COG (WJ–III Gs). Gs marker tests were thus Digit Symbol, Cross Out, and WJ–III Gs. The Gv construct was originally intended to be represented by Block Design, Spatial Relations, and Picture Recognition. However, considering the very low correlations that Picture Recognition had with the other two tests, only the first two tests were included in the EFA. As in Experiment 1, PAF was used because the primary goal was to identify common factors and the number of factors to extract was

Table 4
Descriptive Statistics for Ability Test Scores and
Computer-Based Mouse Response Test Scores

Variable	Response Time (in Milliseconds)			
	M	SD	Minimum	Maximum
Digit symbol	92.8	15.0	59	123
Visual matching[a]	556.0	24.5	507	625
Decision speed[a]	538.8	18.8	496	586
Cross out[a]	524.2	9.0	500	542
Block design	52.0	9.6	28	67
Spatial relations[a]	515.5	7.4	501	533
Picture recognition[a]	511.8	6.8	498	526
Visual auditory learning[a]	504.1	8.8	484	527
APM–SF	7.0	2.9	1	12
Symbol digit	89.4	15.6	62	132
Space code game	62.6	7.6	47	85
Space code game[b]	72.1	13.1	52	102
Guided mouse–grid speed	177.1	25.3	122	256

Note—APM–SF is Raven's Advanced Progressive Matrices–Short Form. [a]Scores provided are Rasch scaled W scores centered on 500. [b]Second administration of Space Code game.

Table 5
Correlation Matrix for Ability Test Scores and Computer-Based Mouse Response Tests Scores

Variable	1	2	3	4	5	6	7	8	9	10
1. Digit symbol	–									
2. Visual matching	.56**	–								
3. Decision speed	.38**	.55**	–							
4. Cross out	.45**	.45**	.53**	–						
5. Block design	.34**	.29*	.32*	.36**	–					
6. Spatial relations	.19	.35**	.07	.25*	.51**	–				
7. Picture recognition	.14	−.02	.39**	.15	.14	−.12	–			
8. Symbol digit	.67**	.35**	.38**	.57**	.28	.31*	.29*	–		
9. Space code game	.54**	.60**	.45**	.48**	.24	.37**	.12	.64**	–	
10. Guided mouse–grid speed	.37**	.45**	.35**	.48**	.14	.18	−.02	.44	.58**	–

* $p < .05$.   ** $p < .01$.

decided by inspection of the scree plot (Cattell, 1966), parallel roots analysis (Montanelli & Humphreys, 1976), and the theoretical issues relevant to the present study.

Inspection of the scree plot indicated one or perhaps two factors. However parallel roots analysis indicated a one-factor solution was appropriate. Extracting just the one factor resulted in a factor with high loadings (minimum .69) for all tests except Block Design and Spatial Relations, which had loadings on this factor of .47 and .43, respectively. Considering this pattern of loadings and the theoretical aim of delineating a Gs and Gv factor, it was decided that a two-factor solution would be appropriate, although the two factors were likely to be quite substantially correlated.

A two-factor PAF solution was submitted to Promax rotation (Kappa = 4) and resulted in two correlated factors ($r = .52$). This solution accounted for 54.5% of the variance. The structure matrix of factor loadings is presented in Table 6. The Pattern Matrix indicated almost identical results.

Using the marker tests as a primary guide, it can be seen that the first factor is strongly loaded by the three traditional Gs markers, and the second factor is strongly loaded by the two Gv tests. This replicates the theoretical measurement framework found in Experiment 1, although the two factors are more highly correlated within this sample. Looking at the loadings for the CBMR tests, both Symbol Digit and the Space Code game have loading patterns of magnitude similar to the PP Gs marker tests. This finding

Table 6
Structure Matrix for Principal Axis Factoring of Ability Test Scores (Gs and Gv) and Computer-Based Mouse Response Test Scores With Promax Rotation (Kappa = 4)

Variable	Processing Speed (Gs)	Visual–Spatial (Gv)
Digit symbol	**.752**	.370
WJ–III processing speed	**.695**	.430
Cross out	**.673**	.430
Block design	.404	**.818**
Spatial relations	.368	**.614**
Symbol digit	**.798**	.380
Space code game	**.784**	.392

Note—Bold numbers indicate loadings >.60. WJ–III processing speed is Woodcock Johnson III Test of Cognitive Abilities processing speed cluster score.

supports the interpretation of both Symbol Digit and the Space Code game as Gs measures. Contrary to results for the prototype Space Code test in Study 1, the Space Code game does not show balanced loadings across both Gs and Gv factors.

Correlations With Composite Measures and Other Ability Tests

To further assess the construct validity (concurrent and discriminant) of the CBMR tests, correlations with composite Gs and Gv scores (as described above) and the two other ability tests, APM–SF and VAL, were calculated. To enable a direct comparison with a PP speeded coding task, the same correlations were calculated for the Digit Symbol test. To avoid the problem of part–whole correlations the Gs composite was calculated using all of the PP Gs markers except for Digit Symbol. These correlations are shown in Table 7. It can be seen that the CBMR tests show a very similar pattern of correlations with the composites and ability tests as Digit Symbol. Guided Mouse–Grid Speed shows the simplest relationship with the Gs composite in that it is not significantly related to Gv or VAL. It also appears that Symbol Digit and the Space Code game may, to some extent, tap an associative memory ability more than does Digit Symbol, as evidenced by these tests having similar but small relationships with the VAL test. Although the CBMR tests do not share a PP methodology with the tests comprising the Gs composite, both of these tests are in fact slightly more strongly correlated with these than Digit Symbol is. All of the tests share a similar relationship with the APM–SF.

GENERAL DISCUSSION

Results across the two experiments considered together suggest that a computer game-like test of Processing Speed (Gs) may provide a viable alternative to traditional PP measures. In Experiment 1, it was found that the simple computerized coding test loaded uniquely on the Gs factor. These results support the previous finding by McPherson and Burns (2005), who also found that this test loaded on a Gs factor independently of a second factor loaded by reaction time tasks. In Experiment 1, it was also found that Space Code, as well as loading on the Gs factor, additionally loaded on the Gv factor. This sug-

Table 7
Correlations Between Computer-Based Mouse Response
Test Scores and Composite Gs and Gv, Raven's Advanced
Progressive Matrices–Short Form (APM–SF) and
Visual-Auditory Learning (VAL) Test Scores

Variable	Gs	Gv	APM–SF	VAL
Digit symbol	.57**	.31*	.38**	.09
Guided mouse–grid speed	.53**	.18	.46**	.07
Symbol digit	.63**	.34**	.45**	.25*
Space code game	.67**	35**	.49**	.24

$^*p < .05.$ $^{**}p < .01.$

gested the possibility that the use of more complex visual stimuli introduces performance variance associated with visual abilities. However, in Experiment 2, this result was not replicated, and, instead, the Space Code game had a loading on the Gv factor similar to the Gs marker tests. It seems unlikely to us that this resulted from the addition of the more dynamic game-like elements to the task, as these did not have any effect on the nature of the visual stimuli. We emphasize that further research is necessary to more fully assess the degree to which Gs tests using game-like stimuli may tap both and Gs- and Gv-related processes.

Despite this issue, the potential of a game-like test of Gs was further evident in the results from Experiment 2. Similar correlations between Gs marker tests and Space Code to those obtained in Experiment 1 were found but using a larger range of marker tests. Both the Space Code game and the simpler coding task showed discriminant validity in that these were not strongly related to a measure of associative learning or delayed recall (Glr), although the relationships were slightly higher than those obtained between Digit Symbol and this measure. This suggests that the tests primarily measure the speed of coding and not the ability to memorize associations between the numbers and spaceships or symbols. Space Code also shows signs of being psychometrically reliable with a test–retest correlation of .84.

Interestingly, all of the computer-based tests had a moderate to strong relationship with Raven's APM–SF. It should also be pointed out that Guided Mouse–Grid Speed showed a substantial relationship with the Gs composite measure and Raven's APM–SF. Thus, this task, which was primarily designed to provide a measure of how quickly people could utilize the mouse response method, may in fact tap cognitive processes more central to traditional abilities tests. The finding that Space Code showed a substantial relationship with Raven's APM–SF, a short-form version of a test considered by many researchers to be central within the spectrum of human cognitive abilities (Carpenter, Just, & Shell, 1990; Prokosch, Yeo, & Miller, 2005), further supports the notion that game-like tests can be developed to supplement traditional cognitive abilities tests.

Although we are optimistic about the potential for developing computer game-like tests of cognitive abilities, a number of important issues remain to be investigated. These include the nature of any possible gender bias effects associated with various game elements. A recent survey of computer-game users suggests that there are differences in the preferences of males and females in regards to types

of games and elements which are most important to them within gaming environments (Wood, Griffiths, Chappell, & Davies, 2004). In Experiment 1, gender and Digit Symbol performance were negatively correlated, replicating the commonly found phenomenon of faster female performance on this test, but gender and Space Code were positively correlated. Further research is necessary to ascertain whether this result is stable, and, if so, what might underlie differential performance on these tasks. Other important issues include whether noisy game-like environments (aurally and visually) affect particular individuals or groups differently and whether different cover stories are more engaging for some people. Nevertheless, we believe that these issues are not seriously detrimental to the development of computer game-like assessment tools and that there are many potential uses for such tools.

As computer technologies continue to progress, it would seem a wasted opportunity for psychological assessment to ignore new possibilities that are created by this ever expanding set of tools. Indeed, it would seem that many researchers are already beginning to venture out of traditional territory with the development of complex simulation assessments (Kroner, Plass, & Leutner, 2005). We believe that computer game-like assessments could similarly provide a useful and interesting addition to the types of assessments traditionally used in abilities research and practice. This process should involve assessing whether computer game-like tests can be designed to measure already identified constructs, as we have attempted to do in the present article, as well as whether there are in fact new constructs tapped by game-like environments.

REFERENCES

Alderton, D. L., Wolfe, J. H., & Larson, G. E. (1997). The ECAT Battery. *Military Psychology*, **9**, 5-37.

Anderson, M. (1992). *Intelligence and development: A cognitive theory*. Oxford: Blackwell.

Bennett, R. E. (1998). *Reinventing assessment: Speculations on the future of large-scale educational testing.* Retrieved September 27, 2006, from www.ets.org/research/researcher/PIC-REINVENT.html.

Berger, A., Jones, L., Rothbart, M. K., & Posner, M. I. (2000). Computerized games to study the development of attention in childhood. *Behavior Research Methods, Instruments, & Computers*, **32**, 297-303.

Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational & Psychological Measurement*, **58**, 382-399.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence-test measures—A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review*, **97**, 404-431.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, **40**, 153-193.

Cattell, R. B. (1963). Theory for fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, **54**, 1-22.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, **1**, 245-276.

Deary, I. J., & Caryl, P. G. (1997). Neuroscience and human intelligence differences. *Trends in Neurosciences*, **20**, 365-371.

Donchin, E. (1995). Video games as research tools: The Space Fortress Game. *Behavior Research Methods, Instruments, & Computers*, **27**, 217-223.

Embretson, S. E. (2003). *The second century of ability testing: Some predictions and speculations* (No. PIC-ANG7): Educational Testing Service: Policy Evaluation and Research Center.

Epstein, J., & Klinkenberg, W. D. (2001). From Eliza to Internet: A brief history of computerized assessment. *Computers in Human Behavior*, **17**, 295-314.

Graham, G., Cook, M., Cohen, M., Phelps, M., & Gerkovich, M. (1985). STAR: A unique embedded performance assessment technique. *Behavior Research Methods, Instruments, & Computers*, **17**, 642-651.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf–Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 53-91). New York: Guilford.

Jackson, D. N., III, Vernon, P. A., & Jackson, D. N. (1993). Dynamic spatial performance and general intelligence. *Intelligence*, **17**, 451-460.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* New York: Praeger.

Jones, M., Dunlap, W. P., & Bilodeau, I. M. (1986). Comparison of video game and conventional test performance. *Simulation & Games*, **17**, 435-446.

Kroner, S., Plass, J. L., & Leutner, D. (2005). Intelligence assessment with computer simulations. *Intelligence*, **33**, 347-368.

McGrew, K., & Flanagan, D. P. (1998). *The intelligence test desk reference* (*ITDR): Gf–Gc cross-battery assessment*. Boston: Allyn & Bacon.

McPherson, J., & Burns, N. R. (2005). A speeded coding task using a computer-based mouse response. *Behavior Research Methods*, **37**, 538-544.

Mead, A. D., & Drasgow, F. (1994). Equivalence of computerized and paper-and-pencil tests: A meta-analysis. *Psychological Bulletin*, **114**, 449-458.

Montanelli, R. G. J., & Humphreys, L. G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika*, **41**, 341-348.

Peterson, N. G., Hough, L. M., Dunnette, M. D., Rosse, R. L., Houston, J. S., & Toquam, J. L. (1990). Project A: Specification of the predictor domain and development of new selection/classification tests. *Personnel Psychology*, **43**, 247-276.

Porter, D. B. (1995). Computer games: Paradigms of opportunity. *Behavior Research Methods, Instruments, & Computers*, **27**, 229-234.

Prokosch, M. D., Yeo, R. A., & Miller, G. F. (2005). Intelligence tests with higher *g*-loadings show higher correlations with body symmetry: Evidence for a general fitness factor mediated by developmental stability. *Intelligence*, **33**, 203-213.

Rabbitt, P., Banerji, N., & Szymanski, A. (1989). Space Fortress as an IQ test? Predictions of learning and of practised performance in a complex interactive video-game. *Acta Psychologica*, **71**, 243-257.

Ree, W. R., & Carretta, T. R. (1998). Computerized testing in the United States Air Force. *International Journal of Selection & Assessment*, **6**, 82-89.

Ryan, E. B. (1994). Memory for Goblins: A computer game for assessing and training working memory skill. *Computers in Human Services*, **11**, 213-217.

Stankov, L. (1998). *g*: A diminutive general. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 19-37). Mahwah, NJ: Erlbaum.

Thurstone, L. L. (1938). The perceptual factor. *Psychometrika*, **3**, 1-17.

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.

Van de Vijver, F. J. R., & Harsveld, M. (1994). The incomplete equivalence of the paper-and-pencil and computerized versions of the General Aptitude Test battery. *Journal of Applied Psychology*, **79**, 852-859.

Washburn, D. A. (2003). The games psychologists play (and the data they provide). *Behavior Research Methods, Instruments, & Computers*, **35**, 185-193.

Wood, R. T. A., Griffiths, M. D., Chappell, D., & Davies, M. N. O. (2004). The structural characteristics of video games: A psychostructural analysis. *Cyberpsychology & Behavior*, **7**, 1-10.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery–Revised*. Allen, TX: DLM Teaching Resources.

Woodcock, R. W., McGrew, K., & Mather, N. (2001). *Woodcock–Johnson III Test of Cognitive Abilities*. Itasca, IL: Riverside Publishing.