The Numerical Distance Effect and Math Achievement:
Assessing the validity of magnitude comparison paradigms

by

Jordan Rozario

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Arts
in
Psychology

Waterloo, Ontario, Canada, 2016

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Abstract

The numerical distance effect (NDE) is the inverse relationship between response times and the distance between two numbers in numerical magnitude comparison tasks. This robust effect has been obtained using multiple magnitude comparison paradigms (MCP). In addition, the size of an individual's NDE has been found to predict mathematical achievement. The present investigation assessed 4 MCP (distance and ratio controlled simultaneous comparison; and primed and non-primed comparison-to-a-standard) for internal reliability, convergent validity, and their ability to predict mathematical achievement and numeracy. Results demonstrate that performance on MCPs correlated with math ability; however, only the NDE in the simultaneous comparison task is uniquely related to math achievement and numeracy.

Acknowledgements

I would like to thank my supervisor, Dr. Jonathan Fugelsang, for his immeasurable patience and insightful guidance, which has been invaluable to me over the past few years. I would also like to thank my family, friends, and colleagues for their support and guidance; specifically, Dr. Evan Risko and Dr. Nathaniel Barr who were instrumental in getting this project up and running. Finally I would like to thank Victoria, whose love and encouragement have supported and motivated me throughout this process.

Table of Contents

Author's Declaration ii

Abstract iii

Acknowledgements iv

Table of Contents v

List of Figures viii

List of Tables ix

Introduction 1

Background and Importance 2

Current Study 5

Hypotheses 5

Methods 7

Participants 7

Procedure 7

Experimental Tasks 8

Results 13

Numerical Magnitude Processing 13

Mathematical Achievement 16

Numeracy 16

Correlation Analysis 17

Regression Analysis 18

Factor Analysis 19

Discussion 21

Replications 21

Internal Reliability 22

Convergent Validity 24

Associations with Math 25

Associations with Numeracy 28

Implications-Grand Summary 30

References 34

Appendices 40

Appendix A 40

Appendix B 42

Appendix C 43

Appendix D 45

Appendix E 46

Appendix F 47

Appendix G 48

Appendix H 49

Appendix I 50

Appendix J 53

Appendix K 54

List of Figures

Figure 1. *Simultaneous Comparison Task* 9

Figure 2. *Comparison-to-a-Standard (Primed) Task* 11

Figure 3. *Distance and Ratio Effects* 15

List of Tables

Table 1. *Numerical Magnitude Descriptives* 13

Table 2. *Distance and Ratio Effect by Slope* 14

Table 3. *Split-Half Reliabilities* 16

Table 4. *Descriptives for the Brief Mathematics Assessment-3* 16

Table 5. *Descriptives for the Numeracy Assessment Questionnaire* 16

Table 6. *Bivariate Correlations* 18

Table 7. *Simultaneous Regression Model Predicting Math Achievement* 19

Table 8. *Varimax Rotated Factor Loadings* 20

Introduction

In the numerical cognition literature, the reliably obtained Numerical Distance Effect (NDE) is prominently reported as an index of an individual's ability to internally process numerical magnitudes. Paradigms typically employed to measure numerical magnitude processing utilize decision-tasks that require participants to assess the relative quantities of numerical stimuli (i.e., is stimulus 'X' quantitatively larger than stimulus 'Y'). The stimuli used in these paradigms can be symbolic numbers such as Arabic numerals and number words or non-symbolic arrays of objects like a grouping of shapes or pictures. The NDE is the inverse relationship between reaction times and numerical distances obtained during these numerical comparison tasks. As the numerical distance between two numerical stimuli increases, the easier it is to distinguish them.

This relation between decision latency and numerical distance is posited to be an emergent property of overlapping mental representations of quantity and numbers (see Feigenson, Dehaene, & Spelke, 2004 for review), which are thought to be the central building block for human mathematical knowledge and performance (Butterworth, 2005). Recently, a competing theory (Monotonic Connection) has sought to explain the NDE as resulting from the preferential weights attributed to numbers as large or small within the context of a given task (Verguts, Fias, & Stevens, 2005). Whether or not the NDE is a result of representational overlap or monotonic connection weights is of both practical and theoretical importance. However, while the results of the present study may add insight to the current debate, the study was not designed to address these questions specifically and any insight gained must be taken with caution. That being said, both theories converge on a key central point: the relation between distance and response time is a result of numerical processing (as opposed to domain general task demands). The purpose of the present study is to assess the validity of this claim.

The current study seeks to evaluate common “bench mark” NDE paradigms in order to test the assumption that the obtained distance effects are a result of the same underlying cognitive process (i.e., numerical magnitude processing), and by extension that they also relate to mathematical ability. These two questions are of critical importance given the widespread use/acceptance of the NDE as a measure of numerical magnitude processing; its sometimes reported connection to mathematical abilities; and the contribution of math to both academic and professional success (Bynner & Parsons, 1997; Parsons & Bynner, 2006).

Background and Importance

The NDE was first reported by Moyer and Landauer (1967), when participants tasked with choosing the larger of two visually presented digits took longer to respond to numerically close pairs than to numerically distant pairs. The NDE has since been found using a variety of other numerical decision tasks: same/different judgements (Sasanguie, Defever, Van den Bussche, & Reynvoet, 2011), and number naming and parity judgements (Reynvoet & Brysbaert, 1999). The most popular decision task is a magnitude comparison -- a task that has spawned several subtypes: simultaneous (side-by-side, overlapping) presentations, sequential presentations, comparisons to a standard, and primed comparisons to a standard. In simultaneous-comparison designs, numerical pairs are presented side-by-side and participants choose either left or right (e.g., Moyer & Landauer, 1967; Bugden & Ansari, 2011; Halberda & Feigenson, 2008; Holloway & Ansari, 2008; Holloway & Ansari, 2009; Reigosa-Crespo et al, 2012; Sasanguie, Van den Bussche, & Reynvoet, 2012; Maloney, Risko, Preston, Ansari, &

Fugelsang, 2010; Sasanguie, De Smedt, Defever, & Reynvoet, 2012a; Sasanguie et al., 2011; Price, Palmer, Battista, & Ansari, 2012); or stimulus arrays overlap and participants choose based on shape or colour (e.g., Halberda, Mazzocco, & Feigenson, 2008; Price et al., 2012; Lindskog, Winman, & Juslin, 2014; Winman, Juslin, Lindskog, Nilsson, & Kerimi, 2014). In a sequential-comparison design, numerical representations are presented one at a time and participants choose either the first or second (e.g., Halberda & Feigenson, 2008; Price et al., 2012; Lindskog et al., 2014). In a comparison to a standard paradigm, the numerical stimuli are presented one-at-a-time and are all compared to a known standard value (Temple & Posner, 1998; Maloney et al., 2010). In a variant of the comparison to a standard paradigm, both sub- and supraliminal primes can been presented prior to the target stimuli resulting in a positive relationship between the numerical distance of the prime from the target and reaction times, known as the Priming Distance Effect (PDE) (e.g., Dehaene et al., 1998; Reynvoet, De Smedt, & Van den Bussche, 2009; Defever, Sasanguie, Gebuis, & Reynvoet, 2011; Koechlin, Naccache, Block, and Dehaene, 1999; Van Opstal, Gevers, De Moor, Verguts, 2008; Sasanguie et al., 2012b; Sasanguie et al., 2011).

The commonality between these magnitude comparison paradigms is the NDE. Independent of task structure and modality (symbolic or non-symbolic), the relation between response time and numerical distance holds true. This has led to the prevalent use of NDEs in numerical cognition literature as an assay for numerical magnitude processing, often with a link to mathematical performance (for review: De Smedt, Noël, Gilmore, & Ansari, 2013). Recent studies (e.g., Maloney et al., 2010; Price et al., 2012; Sasanguie et al., 2011) have raised questions regarding the reliability and validity of the many task variants used to obtain the numerical distance effect. Given the inconsistency that has been reported and the volume of research being published, a critical question addressed in the present study will be to determine if the effects produced by the various paradigms share variance that would be indicative of a systemic commonality -- numerical magnitude processing.

In addition to the above stated goal, the reliability of each task will also be assessed to ensure interpretations are being drawn from appropriately stable effects. Given that numerical processing is thought to be the foundation upon which mathematical knowledge is built (Butterworth, 2005) the extension from numerical processing to mathematical ability is not only logically necessary, but of practical import as math skills and achievement have been linked to success in school and has been associated with future earning potential (Bynner & Parsons, 1997; Parsons & Bynner, 2006). In addition, medical judgement and decision making research has extended interest from general numeracy (knowledge and familiarity with statistics and probability as defined by Schwartz, Woloshin, Black, and Welch (1997)) to include numerical magnitude processing (Lindskog et al., 2014; Winman et al., 2014). Over the past 20 years higher numeracy has been associated with a better understanding of the risks and benefits of medical procedures, which are often explained in terms of percentages (i.e., 80% likelihood of success) or proportions (i.e., 7 out of 8 survival rate). This area of investigation is relatively young and has only investigated non-symbolic magnitude processing in two studies, one of which found a relation to numeracy (Lindskog et al., 2014) and another that did not (Winman et al., 2014).

Based on the examples noted above, it is clear that an ability to recognize and process numerals and quantities efficiently is a necessity in today’s society. Indeed, math and number are so prevalent in society today that it would be nearly impossible to spend an entire day without encountering a single digit or mathematical calculation. It is also clear that the accurate/proper measurement of this basic process will improve our ability to design studies aimed at understanding its source and development, and identifying those in need of assistance/improvement of its accuracy.

Current Study

The main goals of the present investigation are: 1) Replicate the NDEs obtained using common bench-mark comparison tasks; 2) Evaluate the reliability of the obtained NDEs; 3) Assess commonalities between the NDEs in terms of shared variance; 4) Establish the NDE's relation to mathematical ability; and 5) Explore the possible links between numeracy and both math ability and magnitude processing. As the present study focuses on symbolic magnitude processing, the inclusion of a highly used numeracy scale (Schwartz et al., 1997) was for exploratory purposes only. Of note, while the Schwartz et al. (1997) numeracy scale is highly cited, there have been no studies that I am aware of that have investigated its potential relation to actual mathematical performance.

Hypotheses

In line with previous results, it is predicted that performance on each of the numerical magnitude comparison tasks will follow the patterned profile of the Numerical Distance Effect. In the primed manipulation a relation between the prime and target numbers known as the Priming Distance Effect is also expected. Differing patterns of results between studies employing the various task structures have often been explained in terms of domain general processing and response demands that are task related. Having one large population perform each of the comparison tasks in our study, it should stand that any common variance shared between tasks would be related to the numerical processing demands, not task specific ones. Furthermore, if the NDEs obtained from the tasks are truly representative of numerical processing ability, they should be strongly related to mathematical performance. Numeracy is described as fluency with numerical concepts and probability (Schwartz et al., 1997), as such, there should be an association between numeracy and math achievement. An association between numeracy and basic numerical processing is also predicted.

Methods

Participants

Two hundred and twenty-nine undergraduate students from the University of Waterloo (UW) participated in exchange for partial course credit. All participants had normal or corrected to normal vision. Participants received a letter of information (Appendix A) and provided informed consent (Appendix B) before participating. Ethics approval was obtained through a University of Waterloo Research Ethics Committee, and all procedures carried out were within their ethical guidelines. Three participants did not complete all tasks and seven participants were removed because of unusually low task performance (less than 80% accuracy). The data from the remaining 219 participants were subjected to outlier analysis, resulting in the removal of 7 additional participants who had extreme scores on two or more measures (+/- 3 *SD*). The final sample included 212 undergraduates (135 female, M_{age} = 20.23 years, *SD* = 2.83).

Procedure

Participants were tested on an individual basis in a quiet testing room. Testing sessions were approximately 60 minutes in length. All tasks were administered via desktop computer and monitor (Intel ® Core™ 2 Quad, 20 inch LED monitor; see measures section below for task descriptions).

The testing procedure began with the collection of demographic information (age, gender, language, faculty, and major), followed by the six experimental tasks in a counterbalanced order. Using a partial Latin Square, no two participants completed the tasks in the same order. Upon completion of the tasks all participants were provided with verbal and written feedback explaining the purpose of the study (Appendix C).

Experimental Tasks

Simultaneous Comparison Tasks: Distance and Ratio. Both simultaneous comparison tasks required participants to view two simultaneously presented Arabic numbers and decide which was numerically larger. The stimuli were presented side-by-side onscreen and participants indicated their decision via keyboard. The 'Z'(left) and 'M'(right) keys were used for this response and were also colour coded with either an orange or green sticker (respectively) for ease of identification. The procedure for both tasks began with an instruction screen followed by five practice trials. Participants were asked to confirm they understood the task and ask questions before proceeding to the critical trials. Each trial consisted of a fixation-dot for 1000 ms and then stimulus presentation until response (Figure 1).

The stimulus pairs for the DISTANCE manipulation were combinations of the single digits 1-9 (Arial, size 58 font). The pairs were arranged such that each number was paired with every other number resulting in 36 pairs with numerical distances ranging from 1 to 8 (Appendix D). Each pair was presented four times, twice with the larger number on the left and twice with the larger number on the right of the display, resulting in 144 experimental trials. Participants were provided with a self-paced break after the first 72 trials. Response time and accuracy were recorded for each trial.

The stimulus pairs for the RATIO manipulation were generated such that 6 ratios (.25, .33, .5, .66, .75, and .9) were formed. Using numbers ranging from 6 to 53, ten pairs were created for each ratio (Appendix E). Each pair was presented four times, twice with the larger number on the left and twice with the larger number on the right of the display, resulting in 240 experimental trials. Participants were provided with a self-paced break after 120 trials. Response time and accuracy were recorded for each trial.

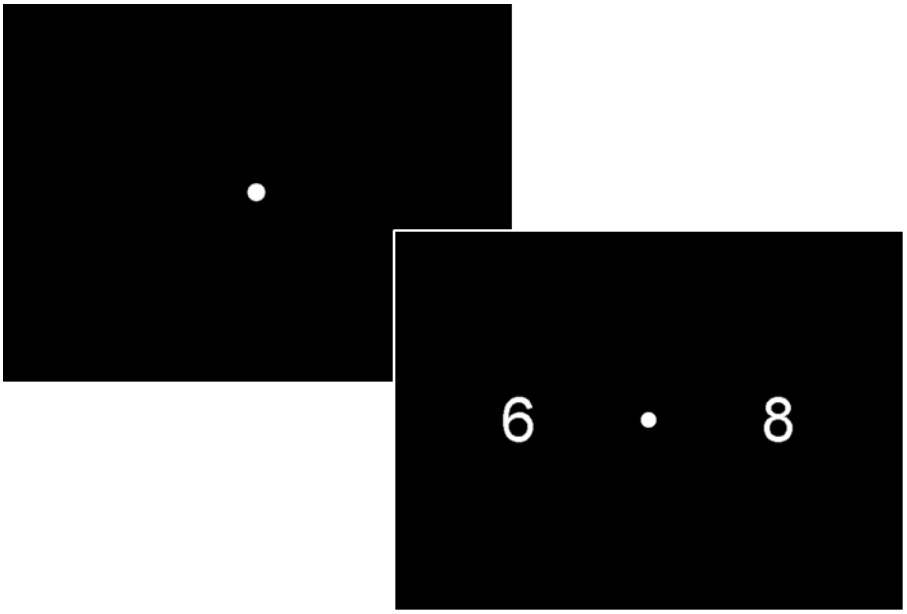


Figure 1. Simultaneous Comparison Task. Pairs of Arabic digits were compared by participants. In the above example there is a distance of 2 and a ratio formed of .75.

Comparison-to-a-Standard Tasks: Non-primed and Primed. Both comparison-to-a-standard tasks required participants to view a single Arabic number and decide if it was numerically larger or smaller than a standard number, 5. Participants indicated their decision via keyboard. The 'Z'(smaller) and 'M'(larger) keys were used for this response and were also colour coded with either an orange or green sticker (respectively) for ease of identification. The procedure for both tasks began with an instruction screen followed by five practice trials. Participants were asked to confirm they understood the task and ask questions before proceeding to the critical trials.

The procedure for the NON-PRIMED comparison-to-a-standard task was a replication of the task reported by Maloney et al. (2010). Trials began with a 500ms fixation dot followed by stimulus presentation until response, following response a blank slide appeared for 500ms, resulting in an ISI of 1000ms. The standard for comparison for this task was also 5. Target numbers were the digits from 1-9 (excluding 5). Each of the targets was presented 20 times for a total of 160 trials. Participants were provided with a self-paced break after 80 trials. Response time and accuracy were recorded for each trial.

The procedure for the PRIMED manipulation was a replication of the task reported by Koechlin et al., (1999, exp. 2a) with two subtle changes. Our procedure used only numerical stimuli (rather than words and numerals) and our ISI consisted of a fixation dot for 1000 ms (rather than 500 ms). The longer ISI duration was used to match the timing of the non-primed comparison-to-a-standard procedure. Trials began with the fixation screen, followed by a forward- and backward-masked prime, the prime and masks were each presented for 66ms (onset synced to the refresh cycle of the display monitor). Participants were not given any special instructions regarding the masks, only that fixation cues would be present prior to the target number. No participants reported seeing the prime numerals. After the masked prime, the target number was presented until response (Figure 2). As in Koechlin et al., (1999, exp. 2a), the target and prime numbers were the single digits 1-9 (excluding 5), each of the eight numbers served as both target and prime an even number of times, resulting in 64 target-prime pairings (Appendix F). Trials with targets 1, 4, 6, and 9 that were primed for a correct response (both prime and target greater- or less-than-5) were considered critical trials. This resulted in 16 of the 64 pairs being critical trial pairs and 48 being distractor trials. Each critical pair was presented 9 times (144 critical trials) while each distractor trial was presented once (48 distractor trials). Thus the experimental session included 192 trials, three-quarters of which were critical. Participants were provided with self-paced breaks after 64 and 128 trials. Response time and accuracy were recorded for each trial.

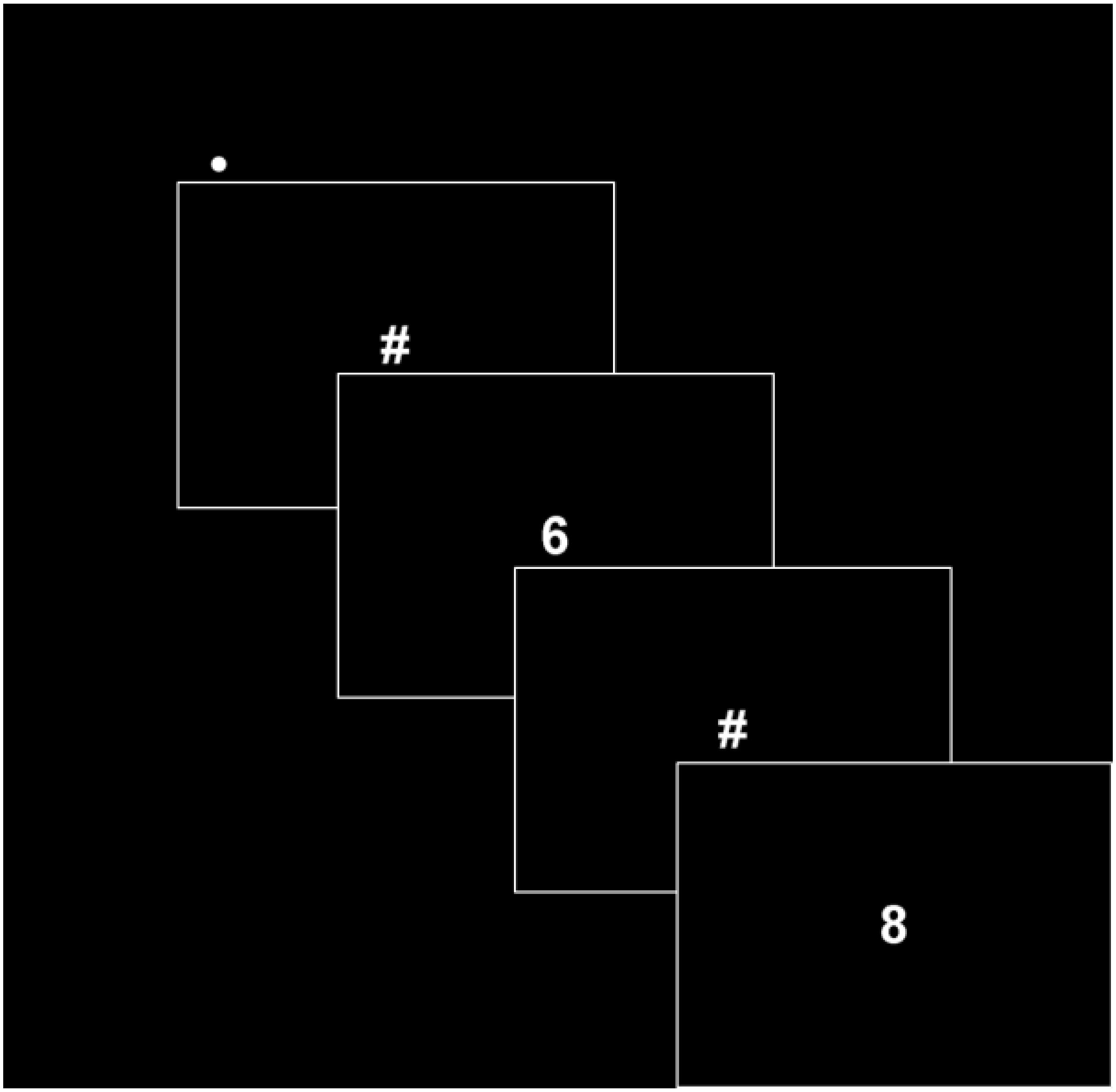


Figure 2. Comparison-to-a-Standard (Primed) Task. Participants decided if a target number was larger or smaller than a standard number, 5. Target numbers were preceded by a forward- and backward-masked prime

Brief Mathematics Assessment-3. Mathematical achievement was measured using the Brief Mathematics Assessment-3 (BMA; Steiner & Ashcraft, 2012; Appendix G). The BMA was administered electronically as well as with paper and pencil. Participants viewed the questions from the BMA, in order, on the computer screen and were asked to input their answers using the keyboard. Concurrently, participants were provided with a paper copy of the BMA and advised they could use the paper for rough work but that they would have to input their final answer via keyboard before proceeding to the next question. The BMA consists of 10 questions that progress in difficulty from multi-digit subtraction to mixed-fraction multiplication including algebra. Participants were encouraged to attempt all questions and were informed of a 10 minute time limit for the task. Following testing, the paper responses and electronic responses were cross-checked by two independent raters, allowing input errors to be corrected. Mean accuracy scores for each participant were calculated and used for analysis.

Numeracy Assessment Questionnaire. Numeracy was assessed with the frequently used Numeracy Assessment Questionnaire introduced by Schwartz, et al. (1997) (Appendix H). The questionnaire was administered via computer as in Pennycook, Cheyne, Barr, Koehler, and Fugelsang (2014) and Barr, Pennycook, Stolz, and Fugelsang (2015). Participants were presented with each of the 3 component questions one at a time on screen and entered their response via keyboard. Mean accuracy scores for each participant were calculated and used for analysis.

Results

As a first step, each Numerical Magnitude Processing task was analysed independently to assess the presence of the associated distance/ratio effects and for internal reliability via split-half reliability calculations. Average scores for math ability and numeracy were calculated. Convergent validity was then assessed via correlation, regression, and factor analysis. A parallel set of analyses were also performed with accuracy data (Appendix I). These results are not presented, nor discussed here as ceiling effects limit the effectiveness of the analyses and the interpretability of the results.

Numerical Magnitude Processing

Analyses were performed using response times and distances or ratio (as applicable) on trials with accurate responses only. Outlier trials (+/- 3 *SD*) were removed on a per subject/distance (or ratio) basis resulting in the removal of at most 1.75% of trials (Table 1). Descriptive statistics for each task including mean response time and accuracy as well as outlier information for each task are displayed in Table 1.

Table 1
Numerical Magnitude Descriptives

Task	Trials	RT Outliers	Mean RT	% Errors
Simultaneous Distance	30528	0.87%	486 (121)	2.4% (2.6)
Simultaneous Ratio	50880	1.29%	555 (156)	4.9% (3.2)
Standard Comparison	33920	1.36%	468 (117)	4.8% (3.5)
Standard Comparison (Primed)	30528	1.75%	489 (125)	3.3% (2.6)

Note. Reaction times are displayed in milliseconds. Std. Deviations are presented in parentheses.

To obtain the size of the NDE, regressions of response time on mean-centred distance/ratio were performed for each participant for each task. The slope of the resulting regression line that best fit each participant's data was used (see: Price et al., 2012). The intercept from the regression line was used for mean response times. Negative DE slopes indicate that shorter response times are associated with larger distances, a positive RE slope indicates that slower response times are associated with larger ratios. The RE slope was reverse coded for consistency in future analyses. To obtain the priming distance effect, regressions of response time on the prime-target distance were performed for each participant (see Dehaene et al., 1998). A positive PDE slope indicates that slower response times are associated with greater prime-target distances, thus the PDE slopes were also reverse coded. Mean response times are plotted as a function of numerical distances or ratios in Figure 3. Single-sample t-tests confirm that the distance effects were obtained on each task, all having negative slopes that are significantly less than 0 (Table 2).

Table 2
Distance and Ratio Effect by Slope

Task	Mean Slope	Std. Deviation	*t*	df	sig. (2-tailed)
Simultaneous Distance	-0.15	0.08	-28.27	211	$p < .001$
Simultaneous Ratio	-0.24	0.08	-43.73	211	$p < .001$
Standard Comparison	-0.17	0.11	-22.94	211	$p < .001$
Standard Comparison (Primed)	-0.05	0.09	-8.24	211	$p < .001$

Note. Mean slopes for each task as well as one-sample t-test results are presented. All slopes were significantly less than 0, indicating that the effects were replicated in all tasks.

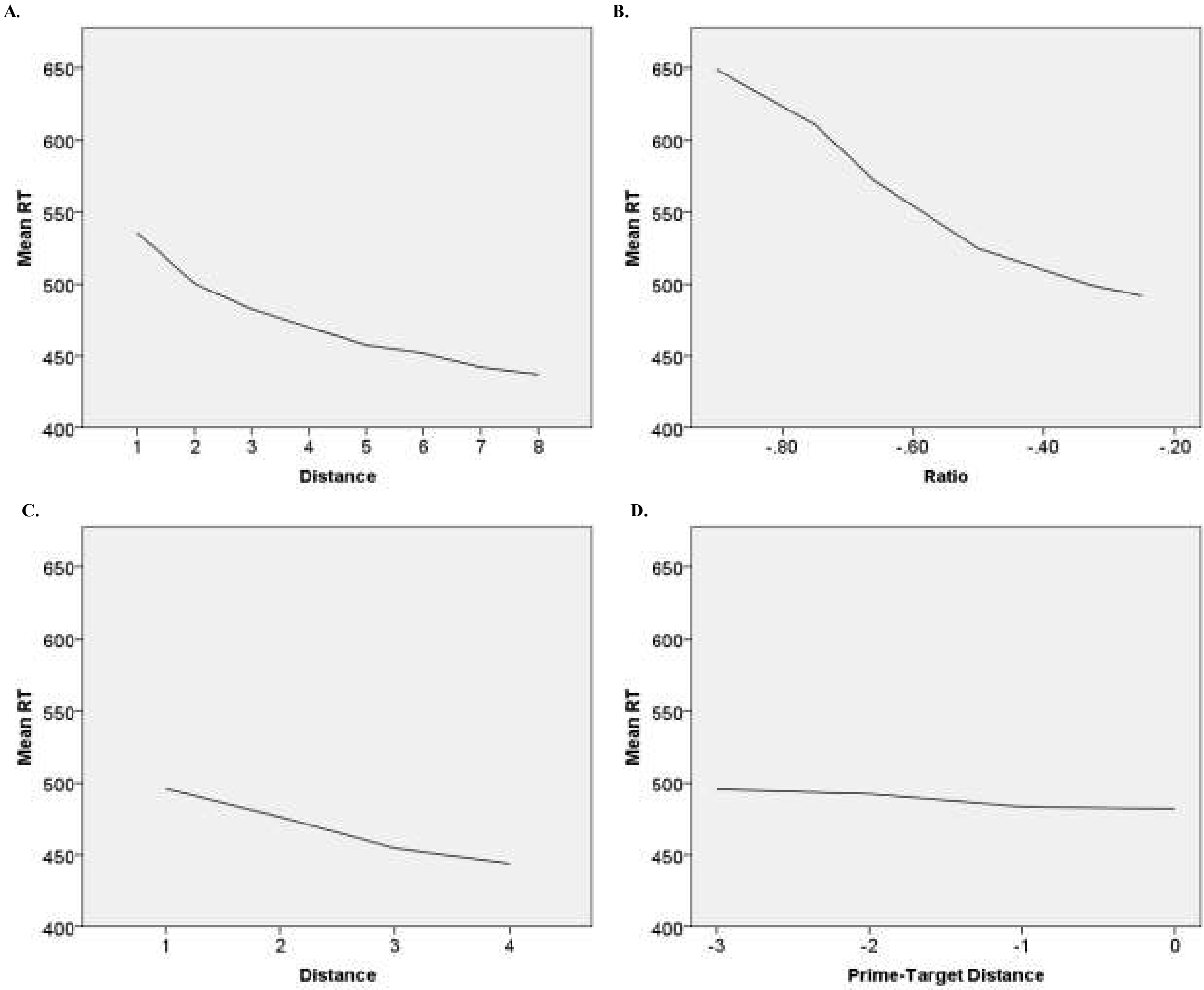


Figure 3. Distance and Ratio Effects. The Numerical Distance Effect was obtained using all four paradigms: (A) Simultaneous Distance; (B) Simultaneous Ratio; (C) Standard Comparison; and (D) Standard Comparison (primed). The Ratio (B) and Primed (D) graphs display the reverse coded data.

Internal reliability estimates were calculated for each of the comparison tasks by calculating the applicable distance effect for each task using the first and second half of trials. These two distance effects were then correlated. The resulting estimates are displayed in Table 3. Fisher *r*-to-Z transformation tests were used to assess statistically significant differences between Pearson correlation coefficients and are presented as applicable in the discussion.

Table 3
Split-Half Reliabilities

Task	Reliability
Simultaneous Distance	.49**
Simultaneous Ratio	.7**
Standard Comparison	.37**
Standard Comparison (Primed)	.24**

Note. ** p < .01

Mathematical Achievement

The mean, standard deviation, and range of raw BMA scores are presented in Table 4.

Table 4
Descriptives for the Brief Mathematics Assessment-3

Mean	Std. Deviation	Minimum	Maximum
7.89	1.92	2	11

Numeracy (probability and numerical concepts)

The mean, standard deviation, and range of raw Numeracy Assessment Questionnaire scores are presented in Table 5.

Table 5
Descriptives for the Numeracy Assessment Questionnaire

Mean	Std. Deviation	Minimum	Maximum
2.3	0.78	0	3

Correlation Analysis:

As a preliminary step, bivariate correlations were calculated for all measures (see Table 6). Mathematical achievement was significantly correlated with all variables except for the distance effect from the standard comparison and the priming distance effect. Higher mathematical achievement is related to better numerical magnitude processing as indexed by the simultaneous presentation tasks as well as mean response times. Better performance on the numeracy task is related to better math achievement, better numerical processing as indexed by the simultaneous presentation tasks, and faster decision times on the simultaneous ratio task. There were 28 possible correlations between decision task slopes and mean RTs, all of which were significantly correlated except for the priming distance effect, which was only correlated with the ratio effect, and the mean response times from the simultaneous distance and standard comparison tasks. Partial correlations were also calculated with age, gender, language, school year, faculty, and major as control variables. Controlling for these individual difference factors did not greatly affect the correlations (Appendix J).

Table 6
Bivariate Correlations

Measure	2	3	4	5	6	7	8	9	10
1. Mathematical Achievement	.27**	.29**	-.24**	.24**	-.23**	.01	-.15*	.00	-0.13*
2. Numeracy	--	.17+	-.12	.25++	-.19++	.06	-.09	-.08	-.09
3. Simultaneous Distance Effect		--	-.58**	.58**	-.53**	.20**	-.33**	.03	-.29**
4. Simultaneous Distance Mean RT			--	-.60**	.76**	-.24**	.60**	-.15*	.54**
5. Simultaneous Ratio Effect				--	-.65**	.26**	-.47**	.12*	-.33**
6. Simultaneous Ratio Mean RT					--	-.28**	.58**	-0.10	.53**
7. Standard Comparison Distance Effect						--	-.58**	0.09	-.43**
8. Standard Comparison Mean RT							--	-.17**	.71**
9. Priming Distance Effect								--	-0.08
10. Standard Comparison (Primed) Mean RT									--

Note. ** *p* < .01, * *p* < .05 (1-tailed); ++ p < .01, + p < .05 (2-tailed)

Regression Analysis:

To further analyse these relations, simultaneous regression analysis was performed with numeracy and all numerical processing tasks entered as predictors of mathematical achievement. The resulting model was significant ($F(9, 202) = 3.94$, $p < .001$), with an R^2 of .15; however, only numeracy and the distance effect obtained using the simultaneous task had significant independent relations to math achievement (see Table 7).

Table 7
Simultaneous Regression Model Predicting Math Achievement

Predictor	B	SE	β	95% C.I. for B	
Numeracy	.15**	.05	0.22	.06	.24
Simultaneous Distance Effect	.44*	.20	0.2	0.05	0.84
Simultaneous Distance Mean RT	0.00	0.0003	-0.08	0.00	0.00
Simultaneous Ratio Effect	0.06	0.21	0.03	-0.34	0.47
Simultaneous Ratio Mean RT	0.00	0.0002	0.01	0.00	0.00
Standard Comparison Distance Effect	-0.16	0.13	-0.1	-0.42	0.09
Standard Comparison Mean RT	0.00	0.0003	-0.06	0.00	0.00
Priming Distance Effect	-0.01	0.13	0.00	-0.26	0.25
Standard Comparison (Primed) Mean RT	0.00	0.0003	-0.02	0.00	0.00

Note. ** $p < .01$, * $p < .05$

Factor Analysis:

Exploratory factor analysis (principal component) was performed to see if the numerical processing measures could be reduced into common factors (Table 8). General assumptions underlying factorability were met. The Kaiser-Meyer-Olkin measure of sampling adequacy was .82, above the minimally acceptable level of .6. Bartlett's test of sphericity was significant ($\chi^2(28) = 791.24$, $p < .001$) thus, the correlation matrix was not an identity matrix. Communalities were all > .5, the numerical magnitude processing measures shared common variance with one another and factor analysis results will account for variance associated with all variables entered.

Three factors were extracted, accounting for 36.1%, 27.7% and 12.8% of the variance. Varimax and oblimin rotations of factor loadings were performed and the resulting orthogonal and oblique solutions were comparable suggesting that the factor structure is truly orthagonal. The varimax rotation solution was therefore used for analysis. The best descriptions of the 3 factors are Simultaneous Comparison, Standard Comparison, and Priming Effect Factors respectively. Both the regression slopes and response times from the simultaneous presentation format tasks grouped together, while the comparison to a standard response times grouped with the slope from the non-primed comparison to a standard distance effect. The priming distance effect had an Eigenvalue of .98 and was a factor on its own. Of the extracted factors, only the Simultaneous Factor was predictive of math achievement (Appendix K).

Table 8
Varimax Rotated Factor Loadings

Component Measure	Simultaneous Factor	Standard Comparison Factor	Priming Effect Factor
Simultaneous Distance Effect	-.82		
Simultaneous Ratio Effect	-.81		
Simultaneous Distance Mean RT	.81		
Simultaneous Ratio Mean RT	.80		
Standard Comparison Distance Effect		-.82	
Standard Comparison Mean RT		.82	
Standard Comparison (Primed) Mean RT		.77	
Priming Distance Effect			.99
Eigenvalues	3.96	1.19	.98
Percentage of total variance	36.06	27.72	12.79

Note. Factor loadings < .4 are suppressed. Rotation converged in 4 iterations.

Discussion

The purpose of the current study was to evaluate the reliability and validity of commonly used symbolic numerical magnitude comparison paradigms (MCP). Reaction time and error rates from MCP tasks typically show the Numerical Distance Effect (NDE), Numerical Ratio Effect (NRE) or the Priming Distance Effect (PDE); collectively referred to as distance effects herein. Previous studies employing these paradigms have posited that these distance effects are related to the human ability to internally process numerical magnitudes (see Feigenson et al., 2004 for review) and have found that performance on MCP can predict mathematical performance and achievement (see De Smedt et al., 2013 for review). The present study was designed to replicate the distance effects of commonly used MCP; evaluate the internal reliability of the distance effects; assess the convergent validity of the distance effects; assess the distance effects' relations to math ability; and evaluate possible relations between math ability, numeracy, and magnitude processing.

Replications

As expected, each MCP task elicited their respective distance effect. The existing literature on the reliability of MCP is very new, with only 3 studies having investigated the issue (Maloney et al., 2010; Price et al., 2012; Sasanguie et al., 2011). Only Maloney et al. (2010) have assessed the reliability of such tasks with symbolic stimuli in an adult population. As such, it was critical that the current study assess both the reliability within tasks and the correlation between tasks to ensure any conclusions drawn were done based on stable and valid effects. Baseline predictions for reliability estimates were that the reliabilities obtained would be similar to those previously reported. The results were in line with this prediction as the split-half reliability coefficients obtained were close to those previously reported. Based on previous findings, a correlation between the L/H 5 and simultaneous paradigms was predicted (Maloney et al., 2010); a prediction of correlations between all other MCPs was driven by their theoretical association to numerical magnitude processing. The results supported this prediction for the non-primed tasks. Furthermore it was predicted that performance on MCP would be related to mathematical performance as previously reported (Lyons & Beilock, 2011; Castronovo & Göbel, 2012; Defever et al., 2011; & Sasanguie et al., 2012b). The results supported the predicted relationships between math and the distance effects from the simultaneous comparison paradigms. The results did not support the predicted relationships between math and the distance effects from the comparison-to-a-standard paradigms. Numeracy was assessed as a possible covariate of math ability, and was also predicted to be associated with numerical processing. In line with these predictions, numeracy scores were related to math ability as well as performance on both simultaneous comparison tasks. I will now discuss each of these analyses and their implications in more detail below.

Internal Reliability

Simultaneous Presentation. Both of the simultaneous comparison tasks were reliable. The ratio-controlled paradigm ($r(210) = .7, p < .01$) was more consistent between blocks ($Z = 3.39, p < .001$). This may be a result of a greater number of trials per block (120 vs. 72), or a property of the selection criteria for stimuli. While not as reliable as the ratio version, the simultaneous distance paradigm was fairly reliable ($r(210) = .49, p < .01$). Previously reported internal reliabilities for simultaneous paradigms were at similar levels. Maloney et al. (2010) performed reliability calculations for two sets of data and found an overall reliability of $r(94) =$

.25 ($p < .05$) when they combined their samples (sample 1: r (46) = .38; sample 2: r(46) = .14). With a much larger sample (n = 212) our result of .49 suggests that the reliability estimates from Maloney et al. (2010) may be under representative of the true reliability of the simultaneous comparison paradigm.

Comparison to a standard (L/H 5). The non-primed L/H 5 comparison to a standard task was reliable (r(210) = .37, $p < .01$). This reliability estimate is significantly greater (Z = 2.55, $p < .01$) than that previously reported by Maloney et al. (2010), who reported a combined reliability of r(94) = .07 ($p > .05$) when they combined their samples (sample 1: r(46) = .06; sample 2: r(46) = .1). Our task design and implementation was a replication of the Maloney et al. (2010) paradigm; the difference in reliability may again be a property of sample size differences.

Priming Distance Effect. The reliability of the primed L/H 5 comparison to a standard task was the lowest of the all tasks (r(210) = .24, $p < .01$). This was the first time that reliability of the masked-prime paradigm was reported, and as such will serve as a baseline for future studies employing the design. Sasanguie et al. (2011) assessed the reliability of the PDE as obtained from adults using an overt respond-to-prime task (participants first respond to a prime and then to the target) and obtained a reliability coefficient of r(45) = .21 ($p = .17$). The two coefficients are not significantly different ($Z = .19$, $p = .42$). The respond-to-prime and masked-prime tasks are theoretically quite different but as an initial starting point to assess reliability, the similarity of the coefficients obtained may indicate that the coefficient obtained by Sasanguie et al. (2011) was representative of the true population coefficient and the non-significant p-value may be an artifact of a small sample size in their study (n = 47). A comparison replication utilizing both paradigms with a sufficiently large sample will help establish the reliability of these tasks.

Reliability Summary. All of the tasks employed to measure numerical magnitude processing were at their core replications of previously used paradigms. With a large sample of adults (212) reliable effects were obtained. The most stable of effects are the Ratio and Distance Effects obtained in the simultaneous comparison paradigms. Lower reliabilities in the comparison to a standard paradigms may be indicative of low task reliability as the obtained reliability coefficients were similar to those previously obtained. When sample sizes are increased and the coefficients do not change, it becomes more likely that the smaller reliabilities are close to the true reliability and the non-significance reported previously was due to a power issue.

Convergent Validity

Response Time. Mean response times (RT) from the MCP tasks were all positively correlated with one another. The strong positive correlations between RTs are indicative of commonality between paradigms. This commonality in RTs could be due to domain general processing speed and it was this assumption that was tested with the factor analysis. The factor analysis revealed that the slopes and RTs from similar presentation formats were more closely associated than just RTs or slopes. The extraction of the format specific factors rather than distance effect or RT factors indicates that domain general processes are influencing performance on these tasks differentially based on the demands created by the task format.

Regression Slope. The slopes of the regression lines for the unprimed tasks were all positively correlated. This indicates that shared variance exists between these MCP tasks. The strongest association was again between variants with similar presentation formats. As evidenced by the factor analysis, there may be some task-specific demands that influence performance on these tasks. The slope from the PDE did not correlate to any of the other slopes. Given that the PDE is posited to be a more pure index for magnitude processing, it was surprising that it was not related to any of the other slopes.

Response Time and Regression Slope. With the exception of the PDE slope, all slopes and RTs from the MCP tasks were significantly correlated.

Convergent Validity Summary. The non-primed MCP demonstrated convergent validity with all slopes and RTs being correlated with one another. The correlations do not indicate whether or not the convergence is due to the domain general processing demands of the tasks or the numerical magnitude processing demands of the tasks. Factor analysis separating the tasks by format is indicative of task specific processing demands. The nuances of the presentation format may be creating task specific demands that may be independent of magnitude processing.

Associations with Math

Simultaneous Distance Effect. The NDE and mean RT from the distance-controlled simultaneous comparison paradigm both correlated with math achievement. The correlation coefficients were the largest for any of the obtained distance effects and mean RTs. Not surprisingly, the NDE from this task was the strongest predictor of math in the regression analysis. The NDE from the simultaneous distance effect paradigm significantly predicts math even when accounting for gender and university program major (appendix J). These results are in line with the previous finding of a correlation between the simultaneous symbolic distance effect and math in adults (Lyons & Beilock, 2011). Using Weber fractions (w) to represent the NDE, Lyons and Beilock (2011) reported a correlation of -.305 (note: smaller w's indicate higher precision leading to a negative correlation). The correlation obtained in our study was .29, and remained strong at .26 even after covariates were controlled (Recall that NDE slopes are negative and larger NDE slopes closer to 0 indicate higher precision leading to a positive correlation). Again, our sample (n = 212) was much larger than that of Lyons and Beilock (2011) (n = 54), lending credence to the assumption that the true correlation between the NDE in this MCP and math achievement is likely around the .3 level.

Simultaneous Ratio Effect. The NRE and mean RT from the ratio-controlled simultaneous comparison paradigm both correlated with math achievement. While the correlation coefficients were almost as large as those obtained in the distance controlled paradigm, they did not prove to be strong independent predictors of math achievement. While the partial correlation for the NRE and mean RT were still significant when controlling for gender and university program major (Appendix J), in the regression analysis the NRE and mean RT from this task did not account for variance in math achievement when the other tasks were included.

Comparison to a Standard. The NDE and mean RT from the non-primed standard comparison paradigm provide puzzling results. The mean RT was correlated with math achievement as predicted and previously reported by Castronovo and Göbel (2012), but the obtained NDE was not significantly associated with math. Castronovo and Göbel (2012) found a strong association between the NDE and math achievement using the WRAT-4, the standardized math assessment that the BMA-3 was built from (Steiner & Ashcraft, 2012). The discrepancy between the results obtained in the current study and those reported by Castronovo and Göbel (2012) are likely due to several important methodological and analytic-procedural differences. In the Castronovo and Göbel (2012) study 73 adult participants performed a lower/higher than sixty-five decision task. This standard comparison in itself is quite different from lower-higher than five (e.g., mappings of large multi-digit numbers are less likely to be as salient as those of single digit numbers), however it has previously been used to obtain NDEs (e.g., Dehaene et al., 1998). The critical difference may be in the type of analyses performed on the data. Castronovo and Göbel (2012) were investigating group differences in numerical processing and math performance and used distance to predict accuracy in their calculation of the regression line rather than predicting RT with distance. While accuracy based NDEs are not unconventional, NDEs based on RT are more widely used (for review see De Smedt et al., 2013). The more curious aspect of the use of accuracy instead of RT is that the authors obtained a significant main effect of distance in their ANOVA analysis of RTs (another widely accepted confirmation of the NDE), but they did not obtain a main effect of group and thus focused their analyses on accuracies for which there was a group difference (Castronovo & Göbel, 2012).

Priming Distance Effect. The PDE did not correlate with math achievement; however, the mean RT from the primed standard comparison paradigm was correlated with math achievement. This correlation did not remain in the regression analysis, however. Previous investigations of the relation between PDEs and math ability have focused on children (Defever et al., 2011; Sasanguie et al., 2012b). Both of these previous studies had large samples (116 & 72, respectively), but while Defever, Sasanguie, Gebuis, and Reynvoet (2011) reported a significant relation between math and the PDE, Sasanguie et al. (2012b) did not. Both studies used the mathematics component of the Flemish Student Monitoring System (Dudal, 2000), both samples included students from kindergarten, grades 1 and 2, and in Defever, Sasanguie, Gebuis, and Reynvoet (2011), grade 6. Neither study reported an effect of grade or age on the PDE. One procedural difference that may account for the difference in the reported relation to math is that Sasanguie et al. (2012b) measured mathematics 1 year post-PDE testing. The inconsistencies in procedure and results between our study of adults and the previously reported studies of children leave the PDE open for further investigation regarding its possible relation to math.

Associations with Math Summary. The NDE obtained from the simultaneous distance-controlled paradigm was the strongest predictor of math achievement. The relation between math and the NDE remained strong, even when controlling for numeracy, gender, and school major (Appendix J). The NRE and mean RTs from the ratio and both standard comparison tasks were initially correlated with math achievement, but the association was reduced significantly by the inclusion of the simultaneous comparison DE and mean RT.

Associations with Numeracy

Magnitude Comparison Paradigms. The investigation of numeracy and magnitude processing was exploratory. As the first study to assess a possible relation between symbolic MCP and numeracy, our predictions were based on the assumption that numeracy, requiring high-level mathematical knowledge, would be less related to magnitude processing than math achievement. Our data supported this prediction as the correlations between numeracy and measures of magnitude processing were of a smaller scale than those between math and magnitude processing. Winman et al. (2014) reported a non-significant correlation between non-symbolic numerical magnitude processing and numeracy (*r*(213) = .1, n.s.), while Lindskog et al. (2014) reported a significant correlation between non-symbolic numerical magnitude processing and numeracy ($r(100) = -.29$, $p < .05$). In our sample, correlation coefficients ranged from .06 to .25. Common to all three studies is a large sample of adults. In theory, numeracy and numerical magnitude processing ability should be stable in the adult population ruling out the possible developmental explanation for the different findings. The differing results may be a result of different numeracy assessments, as the present study used the numeracy scale initial conceived by Schwartz et al. (1997), whereas Lindskog et al. (2014) employed the Berlin Advanced Numeracy Test, and Winman et al. (2014) generated a composite numeracy score using tests that included both the Berlin Advanced Numeracy Test and the Expanded Numeracy test, which includes the items on the Schwartz et al. (1997) scale.

Mathematical Achievement. The present study is the first to assess the association between the Schwartz et al. (1997) numeracy scale and math achievement. Lindskog et al. (2014) assessed arithmetic fluency, but did not include any analysis regarding any possible association with their numeracy measure. As predicted, there was a correlation between math achievement and numeracy, as well as between numeracy and magnitude processing. Furthermore, numeracy was a strong independent predictor of math achievement. Numeracy predicts math achievement even when controlling for numerical magnitude processing ability (as indexed by any of the distance effects or mean RTs), as well as both gender and university major.

Numeracy Summary. Numeracy, knowledge of and familiarity with statistics and probability, is a high-level mathematical knowledge that is related to basic arithmetic and math achievement. The pathway of connections from basic numerical processing through arithmetic to numeracy must be investigated in future studies. If basic numerical processing, that develops in children well before they are introduced to formal probability and statistical learning, can influence the ability to develop higher-level math understanding and ability, steps must be taken to ensure this relationship is well understood and monitored in potential at-risk children such as developmental dyscalculics.

Implications-Grand Summary

The overarching theme of the present investigation was to assess bench-mark numerical magnitude comparison paradigms for their claims that the characteristic Numerical Distance Effect (NDE) produced by the various paradigms is an emergent property of our ability to internally process numerical quantities. To assess this assumption we attempted to link the NDEs to a more complex measure of numerical processing, mathematical performance.

Critically, we successfully obtained the characteristic Numerical Distance Effect in each of the paradigms. With the exception of the Priming Distance Effect (PDE), all of our obtained NDEs were correlated with one another. This initial finding strongly implies that the relation between response time and numerical distance is consistent between task formats and may in fact be due to the numerical nature of the decision tasks. However, given that both task formats can be thought of as forced-choice logical decisions (yes/no in the standard comparison paradigm and right/left in the simultaneous paradigm) their correlation could also be a result of domain general decision processing.

Mathematical ability is highly reliant on both numerical processing and domain general logical processing, so regardless of the underlying factor behind the NDEs, they should have all been related math. The lack of a correlation between the comparison to a standard NDE and math ability is therefore quite puzzling. A reasonable explanation for this seemingly odd outcome becomes apparent when we consider the results of our factor analysis. The extraction of task-format specific factors has two important implications. First, while both formats are logical decision tasks, there is clearly a difference between yes-no confirmatory decisions and left-right multiple-option decisions. Choosing between two options may be a logical decision that is more strongly related to mathematical logic than a more simple yes-no decision. Secondly, when taken together with the bivariate correlations between the NDEs, the separation by format strongly suggests there is still something common between the tasks. This could still be numerical processing as both decision do require the processing of numeric quantities but without a proper control task we cannot rule out general logic processing.

Taken together, this series of results suggests that the common factor between the NDE paradigms might not be numerical processing. Because the NDEs are all correlated without mathematical performance being the connecting factor, it is possible that logical decision processes are the source of the commonalities. We used math to test the theory that the commonality was numerical quantity processing; however, without having controlled for domain general logical decision processing we are unable to rule out the possibility that the logical demands of the tasks are not the connecting factor as math is deeply rooted in logic.

Let us now consider the implications of the regression analysis in which the only NDE to be a unique predictor of math was the simultaneous distance-controlled paradigm. Focusing on both the simultaneous distance-controlled and the simultaneous ratio-controlled comparison paradigms, we can draw some more informative conclusions on the basis of the preceding results. The NDEs from both of the simultaneous comparison tasks correlated with one another implying shared systematic variance, which could be due to the numerical processing or the general decision logic of the simultaneous comparison tasks. Both tasks also correlated with mathematical performance, which again could be due to numerical processing or general logic processing. As the factor analysis grouped tasks by format, we are forced to the assumption that there is task-format based commonality. The ratio-controlled simultaneous comparison task can then be thought of as a parallel version of the distance controlled task wherein the structure of the task is the same, but the content is varied (i.e., the numerical stimuli are different). The tasks are not different in the decision required, they differ specifically in regards to the numerical quantities that are presented and processed.

Again, in the regression analysis, the only predictive distance effect was that obtained in the simultaneous distance-controlled paradigm. When included in the regression, no other distance effects or mean response times were predictive or correlated with math achievement. Because the ratio-controlled version of the simultaneous task can be considered a parallel logical decision task to the distance-controlled task it is reasonable to conclude that the predictive relation between the NDE obtained and mathematical ability is due to numerical processing. By this logic we can also state that the correlation between the ratio effect and math is also likely due to numerical processing; however, the relation is not as strong as the NDE. This is likely due to the stimuli presented during both tasks. The distance-controlled task uses Arabic numbers 1 to 9 presented an equal number of times, whereas the ratio-controlled task uses Arabic numbers ranging from 6 to 53. The processing of larger quantities may be less strongly related to the fundamental building blocks than the core 9 digits of our base 10 system.

It is now evident that Numerical Distance Effects obtained from the simultaneous comparison paradigms are in fact related to our ability to internally process numerical quantities. Comparison to a standard paradigms may also be related to numerical processing but the nature of comparison paradigm provided an inherent confound with general decision logic that we were unable to account for within our current framework. Future studies should include measures of general processing speed, working memory, and intelligence. With more indicators of general logic processing a latent variable approach can be adopted that could generate a model of the true relations between math ability, numerical processing, and general processing.

Future investigations must carefully select an appropriate magnitude comparison paradigm when seeking to assess numerical magnitude processing. Appropriate sample sizes, reliability and validity checks, and general processing measures should be employed in all further studies. This will help ensure that conclusions drawn based on numerical distance effects are not dependent on domain general processing or an unreliable measure.

References

Barr, N., Pennycook, G., Stolz, J.A., & Fugelsang, J.A. (2015) Reasoned connections: a dual-process perspective on creative thought. *Thinking & Reasoning, 21*(1), 61-75.

Bugden, S., & Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition*, *118*(1), 32-44.

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, *46*(1), 3-18.

Bynner, J., & Parsons, S. (1997). *Does numeracy matter? Evidence from the National Child Development Study on the Impact of Poor Numeracy on Adult Life*. London: Basic Skills Agency.

Castronovo, J., & Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PLoS One*, *7*(4): e33832. doi: 10.1371/journal.pone.0033832.

De Smedt, B., Noël, M. P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*(2), 48-55.

Defever, E., Sasanguie, D., Gebuis, T., & Reynvoet, B. (2011). Children's representation of symbolic and nonsymbolic magnitude examined with the priming paradigm. *Journal of Experimental Child Psychology*, *109*(2), 174-186.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., ... & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597-600.

Dudal, P. (2000). LVS-VCLB Leerlingvolgsysteem. Wiskunde Toetsen 1—Basisboek (Student monitoring system. Mathematics Tests 1—manual). Leuven, Belgium: Garant.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307-314.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the" Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457.

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal Number acuity correlate with maths achievement. *Nature*, *455* (7213), 665-668.

Holloway, I. D., & Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Developmental Science*, *11*(5), 644-649.

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1), 17-29.

Koechlin, E., Naccache, L., Block, E., & Dehaene, S. (1999). Primed numbers: Exploring the modularity of numerical representations with masked and unmasked semantic priming. *Journal of Experimental Psychology: Human Perception and Performance, 25*(6), 1882-1905.

Lindskog, M., Winman, A., & Juslin, P. (2014). The association between higher education and approximate number system acuity. *Frontiers in Psychology*, *5:462*. doi: 10.3389/fpsyg.2014.00462

Lyons, I.M., & Beilock, S.L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition, 121*, 256-261.

Maloney, E. A., Risko, E. F., Preston, F., Ansari, D., & Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica*, *134*(2), 154-161.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature,* 215, 1519-1520.

Parsons, S., & Bynner, J. (2006). *Does numeracy matter more?* London: National Research and Development Centre for adult literacy and numeracy.

Pennycook, G., Cheyne, J.A., Barr, N., Koehler, D.J., & Fugelsang, J.A. (2014). The role of analytic thinking in moral judgements and values. *Thinking & Reasoning, 20*(2), 188-214.

Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta psychologica*, *140*(1), 50-57.

Reigosa-Crespo, V., Valdés-Sosa, M., Butterworth, B., Estévez, N., Rodríguez, M., Santos, E., ... & Lage, A. (2012). Basic numerical capacities and prevalence of developmental dyscalculia: The Havana Survey. *Developmental Psychology*, *48*(1), 123.

Reynvoet, B., & Brysbaert, M. (1999). Single-digit and two-digit Arabic numerals address the same semantic number line. *Cognition*, *72*(2), 191-201.

Reynvoet, B., De Smedt, B., & Van den Bussche, E. (2009). Children’s representation of symbolic magnitude: The development of the priming distance effect. *Journal of Experimental Child Psychology*, *103*(4), 480-489.

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012a). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, *30*(2), 344-357.

Sasanguie, D., Defever, E., Van den Bussche, E., & Reynvoet, B. (2011). The reliability of and the relation between non-symbolic numerical distance effects in comparison, same-different judgments and priming. *Acta Psychologica*, *136*(1), 73-80.

Sasanguie, D., Van den Bussche, E., & Reynvoet, B. (2012). Predictors for mathematics achievement? Evidence from a longitudinal study. *Mind, Brain, and Education*, *6*(3), 119-128.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, *127*(11), 966-972.

Steiner, E. T., & Ashcraft, M. H. (2012). Three brief assessments of math achievement. *Behavior Research Methods*, *44*(4), 1101-1107.

Temple, E., & Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-old children and adults. *Proceedings of the National Academy of Sciences*, *95*(13), 7836-7841.

Van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: Comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review*, *15*(2), 419-425.

Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin & Review*, *12*(1), 66-80.

Wilkinson, G. S. (1993). Wide Range Achievement Test 3-Administration Manual. Wilimington, DE: Jastak Associates.

Winman, A., Juslin, P., Lindskog, M., Nilsson, H., & Kerimi, N. (2014). The role of ANS acuity and numeracy for the calibration and the coherence of subjective probability judgments. *Frontiers in Psychology*, *5:851*. doi: 10.3389/fpsyg.2014.00851.

Appendices

Appendix A: Letter of Information

Information & Consent Form

Date: __________

Title of Project:	Measuring Numerical Acuity
Faculty Supervisors:	Dr. Jonathan Fugelsang, Psychology Phone ext: 37197, Email: jafugels@uwaterloo.ca Dr. Evan Risko, Psychology Email: efrisko@uwaterloo.ca
Student Investigators:	Jordan Rozario, Psychology Email: jrozario@uwaterloo.ca Nathaniel Barr, Psychology Email: nbarr@uwaterloo.ca

Study Overview

I am a Master's student in the Department of Psychology at the University of Waterloo conducting research under the supervision of Dr. Jonathan Fugelsang.

You are invited to participate in a study looking at the validity of measures of numerical processing and their relations to mathematical ability.

Past research has shown a relation between numerical processing and mathematical ability. The present research seeks to learn more about the validity of different measures of numerical processing, as well as further examine their relation to mathematical ability.

What You Will Be Asked to Do

This study explores the relation between numerical processing and mathematical ability. It consists of a 60-minute in-lab session. You will complete six tasks in this study; all tasks will be presented on a computer screen (requiring you to respond using the keyboard). The tasks will consist of 4 numerical comparison tasks (i.e., deciding which of two numbers is larger) and 2 mathematical tasks (i.e., solving math equations ranging from simple addition, 1+1, to fractions and algebra).

You do not have to like math or be good at it to do this study.

Participation and Remuneration

Participation in this study is voluntary, and will take approximately 60 minutes of your time. You will receive 1.0 participation credits towards your psychology courses.

Personal Benefits of the Study

The benefits of participation in this study include learning about research in psychology in general and the topic of this study in particular. You will receive additional background information about the study. There are no other personal benefits to participation.

Risks to Participation in the Study

There are no known or anticipated risks associated with participation in this study. Participation is voluntary and you may decline to answer any questions should you find them too difficult or prefer not to answer. You may discontinue participation at any time or refuse to answer any questions without loss of participation credit.

Confidentiality

All information you provide is considered completely confidential; indeed, your name will not be included or in any other way associated, with the data collected in the study. Furthermore, because the interest of this study is in the average scores of the entire group of participants, you will not be identified individually in any way in any written reports of this research. The paper data, with identifying information removed, will be kept for a period of 5 years following publication of the research, after which it will be shredded. The electronic data will be securely stored indefinitely in locked offices in the research laboratory of ***Dr. Jonathan Fugelsang*** in the PAS building to which only researchers associated with this study have access.

Questions and Research Ethics Clearance

If after receiving this letter, you have any questions about this study, or would like additional information to assist you in reaching a decision about participation, please feel free to ask the student investigator or a faculty supervisor listed at the top of this sheet.

I would like to assure you that this study has been reviewed and received ethics clearance through a UW Research Ethics Committee. However, the final decision about participation is yours. If you have any comments or concerns resulting from your participation in this study, please contact Dr. Maureen Nummelin, the Director, Office of Research Ethics, at 1-519-888-4567, Ext. 36005 or maureen.nummelin@uwaterloo.ca.

Thank you for your interest in our research and for your assistance with this project.

Appendix B: Informed Consent Form

Consent of Participant

By signing this consent form, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities.

I have read the information presented in the information letter about a study being conducted by ***Jordan Rozario*** under the supervision of Dr. ***Jonathan Fugelsang*** of the Department of ***Psychology*** at the University of Waterloo. I have had the opportunity to ask any questions related to this study, to receive satisfactory answers to my questions, and any additional details I wanted. I am aware that I may withdraw from the study without loss of participation credit at any time by advising the researchers of this decision.

This project has been reviewed by, and received ethics clearance through a UW Research Ethics Committee. I was informed that if I have any comments or concerns resulting from my participation in this study, I may contact the Director, Office of Research Ethics, at 1-519-888-4567, Ext. 36005.

With full knowledge of all foregoing, I agree, of my own free will, to participate in this study.

____
Print Name

____
Signature of Participant

____
Dated at Waterloo, Ontario

____
Name of Witness

____
Signature of Witness

Appendix C: Debriefing form

Feedback letter:

Project Title: ***Measuring Numerical Acuity***

Student Investigators: Jordan Rozario, Psychology
Email: jrozario@uwaterloo.ca
Nathaniel Barr, Psychology
Email: nbarr@uwaterloo.ca

Faculty Supervisors: Dr. Jonathan Fugelsang, Psychology
Phone ext: 37197, Email: jafugels@uwaterloo.ca
Dr. Evan Risko, Psychology
Email: efrisko@uwaterloo.ca

We appreciate your participation in our study, and thank you for spending the time helping us with our research!

When a person is asked to choose the larger of two numbers they are faster and more accurate when the numerical distance separating the two numbers is relatively large (e.g., 2 vs. 9) compared to when it is relatively small (e.g., 8 vs. 6); participants are also faster when the ratio formed by the number pair is small (e.g., 1 & 4, ratio of .25) compared to when the ratio is relatively large (e.g., 6 & 9, ratio of .66; Moyer & Landauer, 1967).

The numerical distance effect (NDE) and ratio effect (RE) are some of the most robust effects in the study of numerical cognition and form an empirical cornerstone of the theory that numbers are represented spatially. These effects have been replicated several times and have found numerical comparisons to be related to mathematical ability (Holloway & Ansari 2008).

The purpose of this study is to assess whether or not these frequently used measures of numerical processing are indeed all measuring the same thing and to better understand the relationship between these established measures and mathematical ability. Understanding this relationship will provide benefits to educators, students, parents, and curriculum developers, helping to identify deficits and customize future training.

All information you provided is considered completely confidential; indeed, your name will not be included or in any other way associated, with the data collected in the study. Furthermore, because the interest of this study is in the average scores of the entire group of participants, you will not be identified individually in any way in any written reports of this research. Paper records of data collected during this study will be retained for 5 years following publication, in a secured location in PAS, to which only researchers associated with this study have access. After this period, it will be confidentially shredded. Electronic data will be kept indefinitely on a secure location in PAS, to which only researchers associated with this study have access. All identifying information will be removed from the records prior to storage.

This project has been reviewed by, and received ethics clearance through a UW Research Ethics Committee. In the event you have any comments or concerns resulting from your participation in this study, please contact Dr. Maureen Nummelin, the Director, Office of Research Ethics, at 1-519-888-4567, Ext. 36005 or maureen.nummelin@uwaterloo.ca.

If you think of some other questions regarding this study, or to request a summary of the findings, please do not hesitate to contact Jordan Rozario (jrozario@uwaterloo.ca). Preliminary results will be available at the end of the semester.

We really appreciate your participation, and hope that this has been an interesting experience for you.

References (related studies that may be of interest to you):

Holloway, I., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology,* 103(1), 17-29.

Maloney, E.A., Risko, E.F., Preston, F., Ansari, D., & Fugelsang, J. Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica*, 134, 154–161.

Moyer, R.S., & Landauer, T.K. (1967). Time required for judgements of numerical inequality. *Nature, 215,* 1519-1520.

Appendix D: Simultaneous Distance Stimuli Pair Listing

#A	#B	Distance
1	2	1
1	3	2
1	4	3
1	5	4
1	6	5
1	7	6
1	8	7
1	9	8
2	3	1
2	4	2
2	5	3
2	6	4
2	7	5
2	8	6
2	9	7
3	4	1
3	5	2
3	6	3
3	7	4
3	8	5
3	9	6
4	5	1
4	6	2
4	7	3
4	8	4
4	9	5
5	6	1
5	7	2
5	8	3
5	9	4
6	7	1
6	8	2
6	9	3
7	8	1
7	9	2
8	9	1

Appendix E: Simultaneous Ratio Stimuli Pair Listing

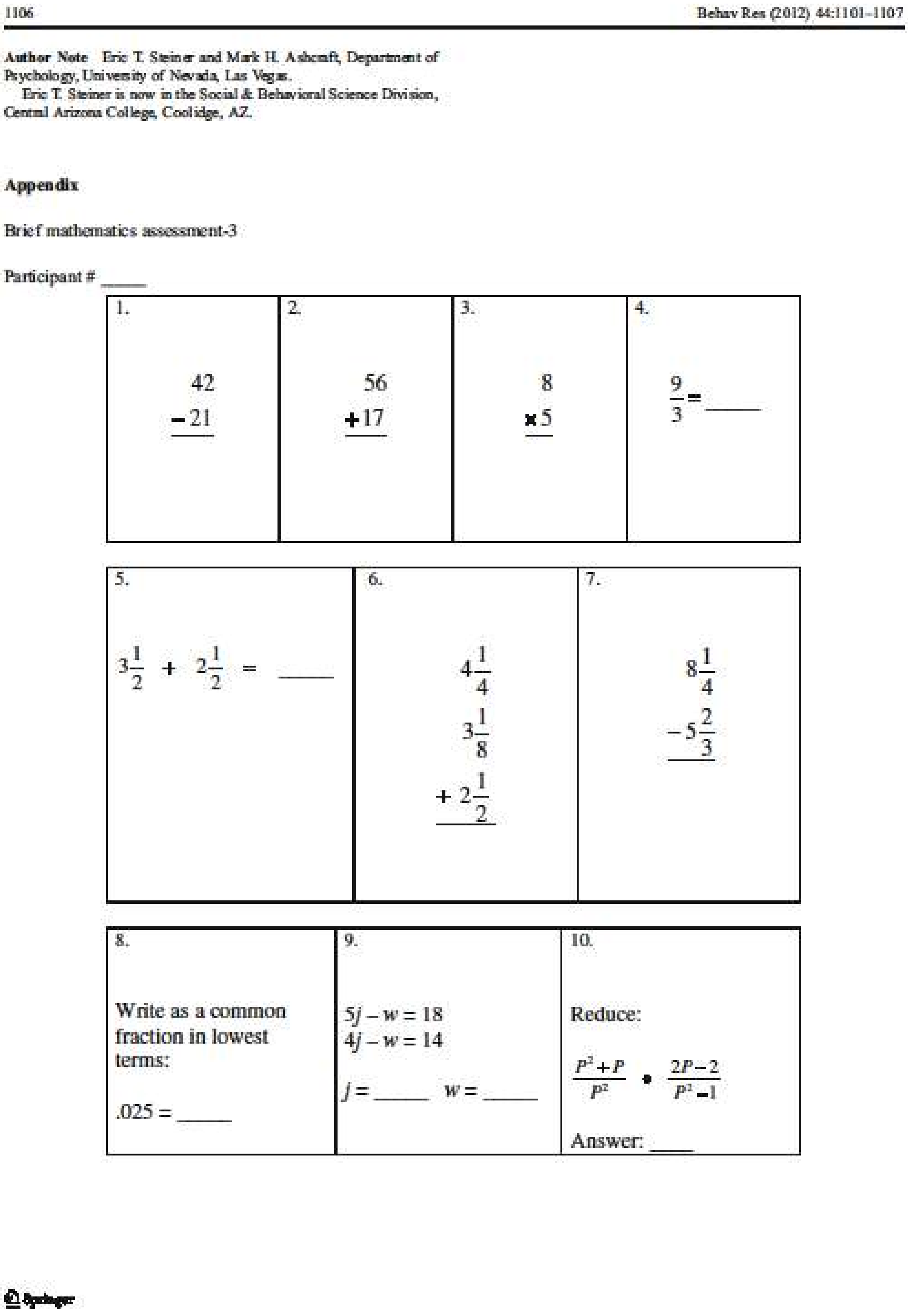
#A	#B	Ratio
6	24	0.25
8	32	0.25
9	36	0.25
10	40	0.25
12	48	0.25
6	18	0.33
8	24	0.33
9	27	0.33
10	30	0.33
12	36	0.33
6	12	0.50
8	16	0.50
9	18	0.50
10	20	0.50
12	24	0.50
6	9	0.66
8	12	0.66
9	14	0.66
10	15	0.66
12	18	0.66
6	8	0.75
8	11	0.75
9	12	0.75
10	13	0.75
12	16	0.75
6	7	0.90
8	9	0.90
9	10	0.90
10	11	0.90
12	13	0.90

#A	#B	Ratio
7	28	0.25
9	36	0.25
10	40	0.25
11	44	0.25
13	52	0.25
8	24	0.33
11	33	0.33
12	36	0.33
13	39	0.33
16	48	0.33
9	18	0.50
12	24	0.50
14	28	0.50
15	30	0.50
18	36	0.50
12	18	0.67
16	24	0.67
18	27	0.67
20	30	0.67
24	36	0.67
18	24	0.75
24	32	0.75
27	36	0.75
30	40	0.75
36	48	0.75
24	27	0.89
32	36	0.89
36	40	0.90
40	44	0.91
48	53	0.91

Appendix F: Standard Comparison (Primed) Stimuli Listing

Prime	Target	Distance
1	1	0
1	2	1
1	9	8
1	8	7
1	7	6
1	6	5
1	4	3
1	3	2
2	2	0
2	1	1
2	3	1
2	9	7
2	8	6
2	7	5
2	6	4
2	4	2
3	3	0
3	2	1
3	1	2
3	4	1
3	9	6
3	8	5
3	7	4
3	6	3
4	4	0
4	3	1
4	2	2
4	1	3
4	9	5
4	8	4
4	7	3
4	6	2
6	6	0
Prime	Target	Distance
6	4	2
6	3	3
6	2	4
6	1	5
6	7	1
6	9	3
6	8	2
7	7	0
7	6	1
7	4	3
7	3	4
7	2	5
7	1	6
7	8	1
7	9	2
8	8	0
8	7	1
8	6	2
8	4	4
8	3	5
8	2	6
8	1	7
8	9	1
9	9	0
9	8	1
9	7	2
9	6	3
9	4	5
9	3	6
9	2	7
9	1	8

Appendix G: BMA-3



1106 Behav Res (2012) 44:1101–1107

Author Note Eric T. Steiner and Mark H. Ashcraft, Department of Psychology, University of Nevada, Las Vegas.

Eric T. Steiner is now in the Social & Behavioral Science Division, Central Arizona College, Coolidge, AZ.

Appendix

Brief mathematics assessment-3

Participant # _____

1. $\begin{array}{r} 42 \\ -21 \\ \hline \end{array}$

2. $\begin{array}{r} 56 \\ +17 \\ \hline \end{array}$

3. $\begin{array}{r} 8 \\ \times 5 \\ \hline \end{array}$

4. $\frac{9}{3} =$ _____

5. $3\frac{1}{2} + 2\frac{1}{2} =$ _____

6. $\begin{array}{r} 4\frac{1}{4} \\ 3\frac{1}{8} \\ +\ 2\frac{1}{2} \\ \hline \end{array}$

7. $\begin{array}{r} 8\frac{1}{4} \\ -5\frac{2}{3} \\ \hline \end{array}$

8. Write as a common fraction in lowest terms:

   .025 = _____

9. $5j - w = 18$
   $4j - w = 14$

   $j =$ _____ $w =$ _____

10. Reduce:

    $\frac{P^2+P}{P^2} \bullet \frac{2P-2}{P^2-1}$

    Answer: ____

Springer

Appendix H: Numeracy Questionnaire

Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips?

In the BIG BUCKS LOTTERY, the chance of winning a $10 prize is 1%. What is your best guess about how many people would win a $10 prize if 1000 people each buy a single ticket to BIG BUCKS?

The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected?

Appendix I: Parallel Analysis Using Accuracy Data

Distance and Ratio Effect by Slope (Accuracy)

Task	Mean Slope	Std. Deviation	*t*	df	sig. (2-tailed)
Simultaneous Distance	0.014	0.01	21.59	201	$p < .001$
Simultaneous Ratio	0.201	0.11	26.37	206	$p < .001$
Standard Comparison	0.026	0.02	19.27	203	$p < .001$
Standard Comparison (Primed)	0.002	0.01	1.98	197	$p < .05$

Note. Mean slopes for each task as well as one-sample t-test results are presented. All slopes were significantly greater than 0, indicating that the effects were replicated in all tasks.

Split-Half Reliabilities (Accuracy)

Task	Reliability
Simultaneous Distance	.17*
Simultaneous Ratio	.48**
Standard Comparison	.23**
Standard Comparison (Primed)	-0.03

Note. ** $p < .01$, * $p < .05$

Bivariate Correlations (Accuracy)

Measure	2	3	4	5	6	7	8	9	10
1. Mathematical Achievement	.27**	-.13*	0.07	-.16*	.16*	-.15*	.28*	.02	.11
2. Numeracy	--	-0.03	0.07	-0.03	0.07	-0.01	0.13	0.05	0.00
3. Simultaneous Distance Effect		--	-.54**	.53**	-.56**	.19**	-.37**	.13*	-.31**
4. Simultaneous Distance Mean ACC			--	-.46**	.71**	-.32**	.48**	-0.07	.40**
5. Simultaneous Ratio Effect				--	-.84**	.35**	-.57**	0.06	-.46**
6. Simultaneous Ratio Mean ACC					--	-.37**	.64**	-0.05	.52**
7. Standard Comparison Distance Effect						--	-.70**	-0.04	-.32**
8. Standard Comparison Mean ACC							--	0.03	.51**
9. Priming Distance Effect								--	-0.04
10. Standard Comparison (Primed) Mean ACC									--

Note. ** *p* < .01, * *p* < .05 (1-tailed); ++ p < .01, + p < .05 (2-tailed)

Simultaneous Regression Model Predicting Math Achievement Using Accuracy

Predictor	B	SE	β	95% C.I. for B	
Numeracy	.15**	.05	0.22	.06	.24
Simultaneous Distance Effect	-1.46	1.64	-0.08	-4.70	1.77
Simultaneous Distance Mean ACC	-1.02	0.72	-0.16	-2.43	0.39
Simultaneous Ratio Effect	0.01	0.22	0.003	-0.43	0.44
Simultaneous Ratio Mean ACC	0.31	0.94	0.06	-1.53	2.16
Standard Comparison Distance Effect	0.17	0.86	0.02	-1.53	1.86
Standard Comparison Mean ACC	1.43*	0.59	0.29	0.26	2.60
Priming Distance Effect	-0.21	0.93	-0.02	-2.04	1.63
Standard Comparison (Primed) Mean ACC	-0.09	0.55	-0.01	-1.18	1.00

Note. ** *p < .01,* * *p* < .05. The model was a significant predictor for math achievement (F(9, 177) = 3.13, p < .01), with an R2 of .14.

Varimax Rotated Factor Loadings (Accuracy)

Component Measure	Simultaneous Factor	Standard Comparison Factor	Priming Effect Factor
Simultaneous Ratio Mean ACC	0.85		
Simultaneous Distance Effect	-0.81		
Simultaneous Ratio Effect	-0.76		
Simultaneous Distance Mean ACC	0.75		
Standard Comparison Distance Effect		-0.91	
Standard Comparison Mean ACC	0.41	0.81	
Standard Comparison (Primed) Mean ACC	0.44	0.49	
Priming Distance Effect			0.99
Eigenvalues	2.90	2.02	1.01
Percentage of total variance	36.15	25.27	12.61

Note. Factor loadings < .4 are suppressed. Rotation converged in 4 iterations. Three factors were extracted and the best descriptions were the same as those from the RT analysis.

Simultaneous Regression Model Predicting Math Achievement Using Extracted ACC Factors

Predictor	B	SE	β	95% C.I. for B	
Numeracy	0.16**	0.05	0.24	0.07	0.25
Simultaneous Factor (ACC)	0.02	0.01	0.10	-0.01	0.04
Standard Comp Factor (ACC)	0.03**	0.01	0.19	0.01	0.06
Priming Effect Factor (ACC)	-0.001	0.01	-0.004	-0.02	0.02

Note. ** *p < .01,* * *p* < .05. The model was a significant predictor for math achievement (F(4, 182) = 5.35, p < .001), with an R2 of .11.

Appendix J: Partial Correlations

Control variables: age, gender, language, school year, faculty, and major

Measure	2	3	4	5	6	7	8	9	10
1. Mathematical Achievement	.22**	.26**	-.20**	.17*	-.16*	.01	-0.13	.06	-.11
2. Numeracy	--	.16+	-.06	.18+	-0.09	.07	-.09	.02	-.06
3. Simultaneous Distance Effect		--	-.57**	.58**	-.54**	.20**	-.32**	.05	-.27**
4. Simultaneous Distance Mean RT			--	-.59**	.75**	-.25**	.60**	-0.13	.52**
5. Simultaneous Ratio Effect				--	-.63**	.27**	-.48**	.18**	-.32**
6. Simultaneous Ratio Mean RT					--	-.29**	.6**	-0.19**	.52**
7. Standard Comparison Distance Effect						--	-.58**	.43**	-.43**
8. Standard Comparison Mean RT							--	-.41**	.71**
9. Priming Distance Effect								--	-0.62**
10. Standard Comparison (Primed) Mean RT									--

Note. ** $p < .01$, * $p < .05$ (1-tailed); ++ $p < .01$, + $p < .05$ (2-tailed)

Appendix K: Extracted Factors Regression Model

Simultaneous Regression Model Predicting Math Achievement Using Extracted RT Factors

Predictor	B	SE	β	95% C.I. for B	
Numeracy	.14**	.05	0.21	.06	0.23
Simultaneous Factor (RT)	-.05**	0.01	-0.26	-0.07	-0.02
Standard Comp Factor (RT)	-0.002	0.01	-0.01	-0.02	0.02
Priming Effect Factor (RT)	-0.00007	0.01	-0.0004	-0.02	0.02

Note. ** *p* < *.01,* * *p* < .05. The model was a significant predictor for math achievement (F(4, 207) = 8.39, p < .001), with an R2 of .14.