Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 86 (2013) 423 – 428

Procedia
Social and Behavioral Sciences

V Congress of Russian Psychological Society

Development and Validation of a Mathematics-number sense Web-based Test Battery

Maria G.Tosto[a*], Tatiana Tikhomirova[b], Eduard Galajinsky[b], Kira Akimova[b], Yulia Kovas[a, b, c]

[a] *Goldsmiths College, University of London, London, SE14 6NW, United Kingdom*
[b] *Tomsk State University, 36 Lenin Prospekt, 634050, Tomsk, Russia*
[c] *King's College, Social, Genetic and Developmental Psychiatry Centre, De Crespigny Park, London, SE5 8AF, United Kingdom*

Abstract

This study aimed to create and validate a web-based battery of tests, age appropriate for 16 year-olds, and designed to assess mathematical skills, general cognitive abilities and number sense. The first stage of the study involved the selection of 11 measures and their administration to a sample of 100 16 year-old students, either in pen and paper format, or on computers. Guided by reliability analyses conducted on the first phase's data, 6 of these tests were selected and implemented online. The new battery revealed to be a reliable tool of assessment showing good internal validity and reliability for all the measures.

Selection and/or peer-review under responsibility of Russian Psychological Society

Keywords: Mathematics; Number sense, Number line; Weber Fraction; Online testing

1. Introduction

The acquisition of mathematical competences is a gradual process requiring the combined effort of cognitive processes including memory, reasoning, IQ, speed of processing [1], [2], [3] and of learned skills, such as reading and writing [4], [5], [6]. However, the relationship between these abilities and mathematics is poorly understood. Recent literature suggests an association between mathematical ability and some cognitive mechanisms that may be unique to mathematics [7], [8]. Abilities underpinning these mechanisms are referred to as "number sense" [9]. Among these abilities, estimation processes have been singled out as predictive of mathematical skills [10], [11].

* Corresponding author. Tel.: +44-207-919-717.
E-mail address: m.tosto@gold.ac.uk (M.G. Tosto)

Selection and/or peer-review under responsibility of Russian Psychological Society
doi:10.1016/j.sbspro.2013.08.591

Processes of non-symbolic estimation do not require number knowledge. When comparing two numerosity sets (grouping of discrete items), it is necessary to extract their numerical information to judge which set is larger; however we can approximate the number without counting all the items. The more the items in display and the smaller the discrepancy between the sets compared, the more difficult it is to discriminate which set has more items. The ability to discriminate according these parameters is indexed by the Weber Fraction. In an example, given two arrays, one with 4 items and one with 6, (ratio 2:3) the Weber Fraction is the ratio derived by dividing the difference of the two set-size (two) by the smallest set-size (four). Non-symbolic estimation is an important process in mathematical cognition; it has been suggested that mathematical disabilities and difficulties can arise from deficits in numerosity processing [3]. Empirical evidence also suggests that this ability improves with development [12], [13]. Moreover, people differ significantly in estimation abilities and that these differences positively correlate with mathematical performance and achievement [8], [14]. Another number sense process is estimation of numerical-symbolic magnitudes. At difference of non-symbolic numerosity estimation, this is based on numerical knowledge. Research suggests that we think about numbers organized on a mental number line in ascending order that follows the direction of writing [7], [15]. Evidence also suggests that numbers are represented in a logarithmic fashion [16]. This is inferred by the fact that we think the distance between small numbers (10 and 20) greater than the distance between larger numbers (90 and 100). In this representation, numbers at the end of the range are closer to one another compared to smaller numbers. Logarithmic mental representation produces less accurate estimates than a linear one. In symbolic estimation tasks, adults are more accurate than children, as they rely on the more efficient linear representation, although a logarithmic one is still present [7], [10], [17].This evidence suggests that symbolic estimation, similarly to non-symbolic estimation, improves with development. Moreover, there is variability in peoples' symbolic estimation skills; these individual differences positively correlate with mathematical skills. [17].

1.1. The present study

The mechanisms driving mathematical acquisition and the etiology of mathematical abilities remain unclear. Further insights could be gained by investigating estimation abilities (number sense) and their etiological overlap with mathematics and with other cognitive abilities that are involved in mathematical acquisition. Further research needs to be planned taking into account the developmental nature of estimation abilities. The purpose of this study was to design and validate a battery of tests, age appropriate for 16 year-olds, in order to assess mathematical abilities, estimation ability and general abilities involved in mathematical acquisition. The aim was to implement the tests in a web-based battery and to create a reliable and practical tool of assessment for the large and widespread UK sample of the Twins Early Development Study (TEDS) at age 16 [18]. In addition to the age of the twins, the choice to carry out investigations at the age of 16 was driven by a gap in the literature as many studies investigating the relationship between mathematics, number sense and cognitive abilities often use very young children. The tests designed for an age-homogeneous sample (16 years) will address issues related to developmental changes in estimation abilities. Analyses will gather more power with data collected in large samples using Internet testing. In addition, a reliable web-based battery will have potential employment for data collection in cross-cultural investigations. The study was designed in two phases: (1) In Phase 1, measures able to assess symbolic and non-symbolic of estimation, general cognitive abilities, and mathematical performance were identified from the current literature. The measures were administered to a sample of 16 year-old students. (2) Phase 2 involved the implementation online of the suitable tests and the administration of this new web-based battery to the same sample of 16 year-olds for reliability and online validation purposes.

2. Methods

2.1. Participants and procedure Phase 1

100 students, with valid data for 98, were recruited from two schools. The sample was composed by 83 females and 15 males, with a mean age of 16.78 (SD = .91). All participants had successfully taken the General Certificate of Secondary Education (GCSE) exam in mathematics. After consent for participating was given by both the students and their legal guardian, students were individually tested, by the experimenter, in a quiet room in the school premises. They were administered 11 tests in various formats; the whole session lasted 40-50 min.

2.2. Measures Phase 1

Estimation skills were assessed with three tests: (1) *Dot Task* was administered via laptop and assessed the number sense ability to estimate large numerosity. This was an adaptation of the task described in details in Halberda et al. [8]. Briefly, students decided whether there were more yellow or blue dots in each of the displays presented in random order. Exposure time was 400ms and response was given by pressing the colour coded keys "F" for more yellow, "J" for more blue. There were 250 trials organised in 5 blocks of 50 and it was possible to pause the test between blocks. A Weber Fraction measure was derived from the accuracy scores as described in Halberda et al.[8]. Cronbach alpha in this assessment was .89, split half of .74. (2) *Number Line* [7] assessed estimation of numerical magnitudes and was in pen and paper format. The task required to mark the position of a given number on a 25 cm long line with the extremes marked 0 and 100. Students estimated 20 numbers given in the same order for all. Score was the error in estimation, measured as the mean absolute difference between the correct location of the number on the line and the point of estimation made. Cronbach alpha was .75, split half .29. The low split half correlation was probably due to the presentation order. The second half of the trials contained more numbers with larger magnitudes compared to the first half. Magnitude comparison in the higher range can be more sensitive to fluctuation in estimation as the numbers are mentally represented more close to one another. The presentation order was modified in the version of the task in Phase 2, where large and small magnitudes were randomly and evenly allocated. (3) *Dot Matching Task* assessed estimation of small and large numerosities. It was online at: http://lab.kctam.com. The 36 trials were displays with a numeral (1 to 9) and arrays of dots presented in pseudo-random order. Students judged whether the number of dots matched the numeral, by pressing "J" for a match and "F" for incongruent trials. Score was total of correct responses. Cronbach alpha was .93, split half .82. Two numerical Stroop tasks assessed numerical information processing and cognitive control and were online in the previous website: (4) *Physical Comparison* required judging the larger in physical size between two numerals irrespectively of the numerical value. (5) *Numerical Comparison* required judging the number larger in magnitude irrespectively of the physical size. Response was given by pressing "J" or "F" if numbers larger in magnitude or in physical size were respectively on the right or left side of the screen. Cognitive abilities were measured with two tasks: (6) *Reaction Time Task* measured speed of processing. It was adapted from Deary et al. [19] and was administered on a laptop. Each of the 40 trials required to press as fast and accurately as possible the key corresponding to the number (1 to 4) appearing on the screen in random order. As the task revealed a ceiling effect on accuracy, efficiency scores (median reaction time divided by proportion of correct responses) were used for the analyses - to correct for the speed/accuracy trade-off. Efficiency scores were normally distributed. Cronbach alpha, calculated on time of response was .96, split half .94. (7) *Corsi Block* assessed non-verbal visuo-spatial short term memory. It was adapted and administered using the Corsi Tapping Block apparatus following the description in Pagulayan et al. [20]. Cronbach alpha was .75 with split half .40. Mathematical ability was assessed with 4 arithmetical tests: (8) *Problem Verification Task* was a timed test assessing mathematical fluency. It was adapted from the description in Murphy & Mazzocco [21] and it was administered on a laptop. Cronbach alpha was .96, split half .94. 3 additional tests assessing mathematical fluency were taken from the website http://lab.kctam.com. These were composed by 20 subtraction and multiplication problems each. Students' GCSEs mathematical scores were also collected.

2.3. Participants and procedure Phase 2

Out of the 100 students recruited in the Phase 1, 75 took part to Phase 2 of the pilot and repeated the tests online. 68 were females and 7 males, with mean age 16.74 (SD = .93). Students were provided with anonimised logins to access the battery at http://www.e-businesssystems.co.uk/teds/ and gave their consent online. The whole battery lasted ~35 min and could be complete in multiple sessions to minimise fatigue; it was designed to be carried out without supervision as all tests had online instructions and practice trials. After completion of the two sessions, students received a £10 shopping voucher for each session completed.

2.4. Measures Phase 2

Analysis of data collected from Phase 1 informed the choice of tests to be implemented online. The 3 online arithmetic tests and the two Stroop tasks revealed a ceiling effect in participants' response with small variation in the distribution of the scores. For this reason and for time constraints the 5 tests were dropped from the battery. A new mathematical test was added. *Understanding Numbers* test assessed mathematical skills according to the standards of the UK National Curriculum. Test items were 27 word problems selected from the nferNelson booklet from the mathematical component "understanding numbers" [22]. It had an adaptive administration with 3 levels of difficulty. Administration was interrupted after three consecutive incorrect responses. Cronbach alpha was .82, split half .61. Some tests from Phase 1 had to be modified for online implementation and time constraints as follow: in *Dot Task* the number of trials was reduced to 150. The subset of stimuli chosen yielded a correlation of .91 with the full length test. The test trials were presented with a fixed sequence to reduce order effects. Cronbach alpha in this assessment was .88, split half .80. Studies show that at 16 years most children rely on linear representations of numerical magnitudes on a mental number line from 0 to 100 [10]. Consistent with these findings, the analysis of the pen and paper *Number Line* data revealed a linear trend in the pattern of estimation. For this reason the range of the estimation of the test in Phase 2 was changed to 0 - 1000. The new version was adapted from a description in Opfer and Sigler [23]. Cronbach alpha in Phase 2 was .87, split half .87. In *Dot Matching Task*, the presentation was fixed the same across participants and the exposure time was reduced to 2 sec to avoid students to count the dots rather than make a judgment response. Cronbach alpha was .85, split half .78. In *Reaction Time* the presentation order was set the same for all participants and the responding keys were changed from letter to number keys. Cronbach alpha on time of response was .94 with split half .91.The number of trials in *Corsi Block* was reduced from 30 to 18 based on an internal validity analysis conducted on data of Phase 1. Cronbach alpha in this assessment was .75, split half .63. In *Problem Verification Task* the confidence rating of response was dropped. Cronbach was .92, split half .87.

2.5. Final version of the measures

The tests described in the above section, were administered to the sample of students in Phase 2. Following this new assessment some tests needed to be modified and optimised for web testing. For other tests the number of trials was reduced further. The description that follows is limited to the modification implemented. *Understanding Numbers*: the total number of items was reduced from 27 to 18. The choice of the items to be dropped was guided by internal validity analysis on the data collected in Phase 2. The difference in internal validity following the item reduction was minimal. Following this change, the discontinue rule applied after two consecutive incorrect responses. *Corsi Block*: the total number of items was reduced from 18 to 12. Internal validity analysis on the data of the Phase 2 of testing showed almost no change in reliability by dropping any of the items. The discontinue rule was modified to account for the reduced numbers of items. *Problem Verification*: the number of items was reduced from 88 to 48. The selection of the items was based on correlation and reliability analysis conducted on the data of Phase 2. The 48 items correlated .97 with the test composed of 88. Cronbach alpha on the 48 items was .86 with split half of .79. The responding keys were also changed for online optimization of the task.

3. Results and discussion Phases 1 and Phase 2

The main aim of this study was to create a valid battery with the power and sensitivity to predict mathematical performance and achievement from measures of number sense abilities in presence of other cognitive abilities. The analyses presented are not aimed at addressing a specific hypothesis, but rather to evaluate the predictive validity and reliability of the measures employed in the newly developed battery. Internal validity of the measures was accurately monitored to make sure that in the translation from different formats, length reduction and implementation on the web, the measures would not lose any of their reliability.

For the 75 students who completed both, one-to-one and online waves of testing, tasks-outliers were identified by creating variables computing the difference of tasks-scores between Phase 1 and Phase 2. For Reaction Time task this variable was created on both Accuracy and Efficiency scores; for all the remaining variables the variable was created only on the accuracy scores. Differences falling outside +/-3 standard deviations were used to identify participants with inconsistent performance between the two Phases of testing - in order to eliminate extreme cases for whom low test-retest stability is likely to stem from unsystematic error. Outlier cases were identified in the frequency tables for each variable and removed on both online and one-to-one tests, for that variable. A total of 25 cases (on different variables) were excluded from the analyses. As Corsi Block and Number Line were in pen and paper format in Phase 1, the comparison of the tests scores between Phase 1 and Phase 2 could have been affected by the change of format. For this reason no variable was created, for these two tasks, the outlier scores were identified if laying outside +/- 3 standard deviations. The analyses were conducted on standardised scores with the exception of Weber Fraction and Reaction Time Efficiency.

Despite the change in administration format, items reduction and parameters adaptation for online implementation, the test-retest correlations were as follow: Problem Verification $r=.60$, $p<.000$, $N=66$; Weber Fraction $r=.43$, $p<.000$, $N=67$; Number Line $r=.30$, $p<.05$, $N=64$; Corsi Block $r=.57$, $p<.000$, $N=61$; Reaction Time Efficiency $r=.41$, $p<.000$, $N=64$. For the least modified tests the correlations were substantial. Only the Dot Matching task had a non-significant re-test coefficient. It has to be noted that during Phase 1, when the task was downloaded, students experienced problems with the stimuli-images. This was probably due to internet bandwidth issues and the image size of the stimuli. Other tasks downloaded from the same website were not affected as the stimuli were only text based. This technical problem was completely resolved in Phase 2.

It was further confirmed that the measures employed behaved consistently with findings in the existing literature. In Phase 1 the two number sense measures Number Line and the Weber Fraction scores significantly correlated with each other (.23, $p<.05$). In addition Weber Fraction correlated with Problem Verification (.33, $p<.01$) but not with GCSEs. Number Line did not correlate with any mathematics measure. Reaction Time Efficiency correlated with both mathematical variables (.27, $p<.01$ with GCSEs and .21, $p<.05$ with Problem verification), Corsi Block correlated with GCSEs scores only (.30, $p<.01$). The different pattern of correlation of number sense and other cognitive ability with mathematics is in line with previous findings [24]. Also, given the small sample size perhaps this pilot did not have enough power to detect a correlation between Number Line and mathematics; alternatively the estimation 0 to 100 used for this assessment was not suitable for 16 year-olds. In Phase 2 the two number sense measure no longer correlated with each other; however Number Line significantly correlated with GCSE score (.35, $p<.01$), but not with Problem Verification or Understanding Numbers. Similarly, Weber Fraction correlated only with one mathematical variable, Problem Verification (.31, $p<.05$). Reaction Time Efficiency correlated with the GCSE scores (.37, $p<.01$) and Problem Verification (.26, $p<.05$) showing again the strength of the association between mathematics and general cognitive ability. Corsi Block correlated only with Problem Verification (.28, $p<.05$).

Overall, this pilot study was successful in selecting and validating measures for a new bespoke tool for the assessment of mathematics, general abilities and number sense skills. The measures implemented online showed reliability and the ability to capture different aspects of the mathematical construct.

Acknowledgements

We are grateful for the support of the grant from the Government of the Russian Federation (11.G34.31.003) and the Goldsmiths College Small Grant.

References

[1] Geary DC, Wiley JG. Cognitive addition: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychol Aging*, 1991. 6(3): 474.
[2] Bull R, Johnston RS. Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *J Exp Child Psychol*, 1997. 65(1): 1-24.
[3] Butterworth B, *Developmental dyscalculia*, in Handbook of mathematical cognition J.I.D. Campbell, Editor. 2005, Psychology Press: Hove. p. 455-467.
[4] Geary DC. Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychol Bull*, 1993. 114(2): 345-362.
[5] Dirks E, Spyer G, van Lieshout EC, de Sonneville L. Prevalence of combined reading and arithmetic disabilities. *J Learn Disabil*, 2008. 41(5): 460-473.
[6] Vukovic RK, Lesaux NK, Siegel LS. The mathematics skills of children with reading difficulties. *Learn Individ Differ*, 2010. 20(6): 639-643.
[7] Siegler RS, Opfer JE. The Development of Numerical Estimation Evidence for Multiple Representations of Numerical Quantity. *Psychol Sci*, 2003. 14(3): 237-250.
[8] Halberda J, Mazzocco MMM, Feigenson L. Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, 2008. 455(7213): 665-668.
[9] Dehaene S. *The number sense*. New York: Oxford University Press; 1997.
[10] Siegler RS, Booth JL. Development of numerical estimation in young children. *Child Dev*, 2004. 75(2): 428-444.
[11] Lourenco SF, Bonny JW, Fernandez EP, Rao S. Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *PNAS*, 2012. 109(46): 18737-18742.
[12] Xu F, Spelke ES. Large number discrimination in 6-month-old infants. *Cognition*, 2000. 74(1): B1-B11.
[13] Feigenson L, Dehaene S, Spelke E. Core systems of number. *Trends Cogn Sci*, 2004. 8(7): 307-314.
[14] Halberda J, Ly R, Wilmer JB, Naiman DQ, Germine L. Number sense across the lifespan as revealed by a massive Internet-based sample. *PNAS*, 2012. 109(28): 11116-11120.
[15] Dehaene S, Bossini S, Giraux P. The mental representation of parity and number magnitude. *J Exp Psychol Gen*, 1993. 122(3): 371.
[16] Dehaene S. The neural basis of the Weber–Fechner law: A logarithmic mental number line. *Trends Cogn Sci*, 2003. 7(4): 145-147.
[17] Booth JL, Siegler RS. Developmental and individual differences in pure numerical estimation. *Dev Psychol*, 2006. 42(1): 189.
[18] Haworth CM, Davis OS, Plomin R. Twins Early Development Study (TEDS): A Genetically Sensitive Investigation of Cognitive and Behavioral Development From Childhood to Young Adulthood. *Twin Res Hum Genet*, 2012: 1-9.
[19] Deary IJ, Der G, Ford G. Reaction times and intelligence differences: A population-based cohort study. *Intelligence*, 2001. 29(5): 389-399.
[20] Pagulayan KF, Busch RM, Medina KL, Bartok JA, Krikorian R. Developmental normative data for the Corsi Block-tapping task. *J Clin Exp Neuropsyc*, 2006. 28(6): 1043-1052.
[21] Murphy MM, Mazzocco MM. Mathematics learning disabilities in girls with fragile X or Turner syndrome during late elementary school. *J Learn Disabil*, 2008. 41(1): 29-46.
[22] nferNelson. *Maths 5–14 series*. London: nferNelson Publishing Company Ltd; 2001.
[23] Opfer JE, Siegler RS. Representational change and children’s numerical estimation. *Cognitive Psychol*, 2007. 55(3): 169-195.
[24] Fuchs LS, Geary DC, Compton DL, Fuchs D, Hamlett CL, Seethaler PM, Bryant JD, Schatschneider C. Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Dev Psychol*, 2010. 46(6): 1731.