

## Learning in colour: children with grapheme-colour synaesthesia show cognitive benefits in vocabulary and self-evaluated reading

Article (Accepted Version)

Smees, Rebecca, Hughes, James, Carmichael, Duncan A and Simner, Julia (2019) Learning in colour: children with grapheme-colour synaesthesia show cognitive benefits in vocabulary and self-evaluated reading. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374 (1787). a20180348. ISSN 0962-8436

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/95444/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Learning in colour: Children with grapheme-colour synaesthesia show cognitive benefits in vocabulary and self-evaluated reading**

R Smees<sup>a</sup>, JEA Hughes<sup>a</sup>, DA Carmichael<sup>b</sup> and J Simner<sup>a</sup>

*<sup>a</sup>School of Psychology, Pevensey Building, University of Sussex. BN1 9QJ. UK.*

*<sup>b</sup>School of Applied Sciences, Edinburgh Napier University, EH11 4BN. UK*

*Correspondence: r.smees@sussex.ac.uk, School of Psychology, University of Sussex. BN1 9QJ.*

**Acknowledgements:** The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. [617678]. We are grateful to Julia Aruacz and Alex Hurst for their help in data entry. We also wish to express our gratitude to the children, and families who contributed to this study.

**Author contributions:** RS and JS designed the study, interpreted the results and wrote the paper. RS and JEAH ran the data analysis.

## **Abstract**

Cognitive benefits associated with grapheme-colour synaesthesia in adults are well documented, but far less is known about whether such benefits might arise in synaesthetes as children. One previous study on a very small group of randomly sampled child synaesthetes found cognitive benefits in short-term memory and processing speed (the ability to quickly scan an array of images and discriminate between them), but was inconclusive for a test of receptive vocabulary. Using a stratified population sample (*Growing UP in Scotland* Project, Scotland. Scottish Executive., 2007), we investigated the performance of a large cohort of child grapheme-colour synaesthetes using four literacy measures taken at age 10 years. These were three verbal comprehension measures (expressive vocabulary, receptive vocabulary, sentence comprehension), and one measure of academic self-concept in reading (plus one measure of academic self-concept in numeracy as a comparison). After controlling for demographic differences between groups, synaesthetes showed significantly enhanced performance for expressive and receptive vocabulary compared to their peers, but no benefits in sentence comprehension. Child synaesthetes also reported significantly higher academic self-concept for reading, but not for numeracy. Finally, we found that synaesthetes made significantly more progress than controls across the primary school years, although they began school with no *a priori* advantage. Our study provides powerful new evidence that children with grapheme-colour synaesthetes show vocabulary and literacy differences, which we contextualise within a theory of synaesthetic development.

**Key words:** Grapheme-colour, Synaesthesia, Expressive Vocabulary, Receptive Vocabulary, Sentence comprehension, Literacy, Academic self-concept,

**Acknowledgements:** The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. [617678]. We are grateful to Julia Aruacz and Alex Hurst for their help in data entry. We also wish to express our gratitude to the children, and families who contributed to this study.

## **Introduction**

Grapheme-colour synaesthesia is an unusual neurodevelopmental trait in which letters, digits, or whole words are experienced as having distinct and automatic colours (e.g., Simner, Glover, & Mowat, 2006). For example, the number 5 may be experienced as red, or the letter S experienced as yellow. Although commonalities in colours-for-graphemes can be seen across very large numbers of synaesthetes (e.g., A is coloured red more often than chance would predict; (Rich, Bradshaw, & Mattingley, 2005; Simner et al., 2005)), on an individual basis most synaesthetes have colour palettes that are relatively idiosyncratic. Studies show that grapheme-colour synaesthesia develops across the life-span, emerging some time during early to mid-childhood (Simner, Harrold, Creed, Monro, & Foulkes, 2009) and then showing age-related declines in old age (Meier, Rothen, & Walter, 2014; Simner, Ipser, Smees, & Alvarez, 2017). Virtually no research whatsoever has been carried out for synaesthetes at either end of the lifespan, and here we focus on grapheme-colour synaesthesia in children up to the age of 12 years.

One important question in the study of childhood synaesthesia is how it might impact on the cognitive or educational abilities of child synaesthetes. This question has received scant attention in the developmental literature, but there are clues from adult studies that synaesthetes may enjoy certain cognitive benefits. In memory for instance, adult synaesthetes have a number of advantages (for review see Meier & Rothen, 2013; Rothen, Meier, & Ward, 2012), performing better than controls in recalling word lists (Gross, Nearing, Caldwell-Harris, & Cronin-Golomb, 2011; Radvansky, Gibson, & McNeerney, 2011; Yaro & Ward, 2007) and word-pair associations (Gross et al., 2011; Rothen & Meier, 2010) for example. Additional advantages for adult synaesthetes also come in creativity and perceptual processing: Ward, Thompson-Lake, Ely, & Kaminski (2008) showed superior performance by grapheme-colour synaesthetes in the *Remote Associates Test* (which measures convergent creative thinking but also relates to verbal ability; Lee, Huggins, & Theriault, 2014; Mednick, 1968) and adult synaesthetes have benefits, too, in colour processing (e.g., recognition and memory for colours; Terhune, Wudarczyk, Kochuparampil, & Cohen Kadosh, 2013; Yaro & Ward, 2007) and visual search (Brang & Ramachandran, 2008; Ramachandran & Hubbard, 2001).

The reason for these advantages is somewhat unclear. One model, relating to *Dual-coding Theory*, suggests that grapheme-colour synaesthesia perhaps makes graphemes more robust in memory because they are additionally encoded with colours (e.g. Radvansky et al., 2011). Support for the dual-coding theory comes from observing that grapheme-colour synaesthetes show advantages in manipulating graphemes in particular (i.e., letters, numbers or words). And this has been found, at least in adults (Chun & Hupé, 2016; Radvansky et al., 2011; Rothen et al., 2012; Yaro & Ward, 2007). However, many other studies provide evidence that grapheme-

colour synaesthesia confers yet *broader* advantages, for example, for visual stimuli such as faces or scenes (Gross et al., 2011; Pritchard, Rothen, Coolbear, & Ward, 2013; Rothen & Meier, 2010; Ward, Hovard, Jones, & Rothen, 2013). The evidence for wider cognitive benefits in grapheme-colour synaesthesia suggest a more global advantage, perhaps through a broader type of enhanced perceptual or structural organisation (Hänggi, Wotruba, & Jäncke, 2011; Ramachandran & Azoulay, 2006; Simner & Bain, 2017), or even differences in “cognitive processing style“ (Meier & Rothen, 2013a). These broader advantages might be mirrored in the relatively global neurological differences that have been found in the brains of synaesthetes, in addition to more localised enhancements (e.g., in colour/ grapheme areas; Jäncke, Beeli, Eulig, & Hänggi, 2009; Rouw & Scholte, 2007). However, evidence of broad advantages does not negate the possibility of dual-coding per se, simply because more than one mechanism might be working in parallel. At the very least, people with grapheme-colour synaesthetes might be expected to show benefits at the level of graphemes and words, and so here we investigate their verbal abilities more closely.

In looking closely at the advantages enjoyed by synaesthetes, an important question is when these differences might emerge. In terms of benefits for synaesthetes as children, only two studies have broached this question. Green and Goswami (2008) tested grapheme-colour synaesthetes 7-15 years of age and found a cognitive *disadvantage* (rather than advantage) in that synaesthetes struggled to remember certain types of coloured digits. Specifically, when children recalled digit matrices, synaesthetes performed worse than controls if digits were coloured incongruently to their synaesthesia (e.g., if the number 5 was presented in red font, but happened to be synaesthetically blue for the child-synaesthete). Later, Simner and Bain (2018) analysed additional data gathered by Green and Goswami and found, conversely, what appeared to be cognitive advantages. The Green and Goswami cohort of synaesthetic children performed significantly better than population norms in a test of receptive vocabulary (*British*

*Picture Vocabulary Scale*; BPVS; e.g. Dunn, Dunn, Whetton, & Pintilie, 1982), and a test of perceptual reasoning (The Blocks task of the *Wechsler Intelligence Scales for Children*; WISC; in which children arranged blocks to match a target pattern; Wechsler, 2004). Importantly however, the recruitment methods of Green and Goswami could potentially have encouraged high performing children irrespective of synaesthesia (see Simner & Bain, 2018 for discussion). It is likely Green and Goswami recruited by parental referral (i.e., parents referred their synaesthetic children in response to a study advert, or some other call for participants). We have argued elsewhere that children referred by their parents are a priori likely to be high performers in cognitive tasks simply because they come from families where parents are motivated to engage in scientific research. Such children may therefore be non-representative of what we might expect from average synaesthetic children, randomly sampled.

To overcome this problem, Simner and Bain (2018) recruited randomly-sampled child synaesthetes by screening very many hundreds of children from local primary schools in Edinburgh to identify the small percentage of grapheme-colour synaesthetes among them (Simner & Bain, 2013; Simner et al., 2009). When these child synaesthetes were aged 10-11 years, they were assessed in cognitive tests for receptive vocabulary (BPVS; Dunn, Dunn, Whetton, & Pintilie, 1982), speed-of-processing (WISC Cancellation task; Wechsler, 2004), and letter-span (i.e., working memory recall of letters; Simner & Bain, 2018). Simner and Bain found that child synaesthetes performed significantly better than population norms in the speed-of-processing task, and trended towards significance in one version of the letter-span task (at  $p=.06$ ). Additionally, in the BPVS, 80% of child synaesthetes scored higher than expected norms although this failed to reach significance, perhaps because the sample size was very small. The rarity of synaesthesia meant that their screening of more than 600 children had delivered only a handful of child synaesthetes, and cognitive tests were therefore carried out

on  $n=5$  synaesthetes. When we evaluate the power of that study (see Supplementary Information, SI) we find that this sample size could only have identified significant differences if effect sizes were very large (Cohen's  $d > .1.0$ ; see SI).

In summary, one child study has shown superiority in cognitive tasks (receptive vocabulary, perceptual reasoning; Green & Goswami, 2008) but with parent-referred samples that may be superior irrespective of synaesthesia. And another child study has tested randomly-sampled child synaesthetes (Simner & Bain, 2018) but only in very small numbers meaning that some findings were marginal or non-significant (e.g., in receptive vocabulary). Here we rectify both problems by testing the cognitive abilities of the largest ever group of child grapheme-colour synaesthetes ( $> n50$ ), recruited in such a way that any benefits can be directly tied to synaesthesia rather than sampling methods. Our interests here lie particularly in benefits in language and literacy, and in specific lexicon knowledge (i.e., vocabulary), given the ambiguous findings for receptive vocabulary (BPVS) in Green and Goswami (2008) and Simner and Bain (2018).

In terms of what we might expect from children, we can consider related findings in synaesthetic adults. One study has shown a moderate positive vocabulary advantage for a heterogeneous group of adults who had one or more of eight different types of synaesthesia (e.g., coloured letters, sounds, tastes; Chun & Hupé, 2016)). These participants were given a subtask of the *Verbal Comprehension* section of the *Wechsler Adult Intelligence Scale (WAIS-III)*; (Wechsler, 1997) which requires participants to give verbal descriptions for vocabulary words. Chun and Hupé found a significant advantage for synaesthetes over non-synaesthetes but their effect sizes (Cohen's  $d$ ) were accompanied by a large degree of uncertainty, as



evidenced by very large 95% confidence intervals that spanned 0 (no effect) and 1. This means that the true difference between synaesthetes and controls within the wider population could either be large or nothing at all. Chun and Hupé (2016) also studied a range of different types of synaesthetes, and used a test of vocabulary that involves elements of both receptive and expressive knowledge (understanding and producing words). Here we re-explore the important issue of vocabulary knowledge raised by Chun and Hupé, but within a single type of synaesthesia (grapheme-colour synaesthetes) and with vocabulary measures that test expressive and receptive knowledge separately. Importantly, we do so while also asking whether advantages reach back into childhood. We test a large sample child synaesthetes (n51), thereby overcoming limitations of smaller samples previously (Simner & Bain, 2017). We also avoided the self-referral confounds in earlier child investigations by screening for synaesthesia rather than relying on parental-referral (and finally, ensuring that both synaesthetes and controls were drawn from the same cohort of children).

Our cohort was taken from the *Growing Up in Scotland* study (GUS; Scottish Executive., 2007) a large longitudinal investigation funded by the Scottish government to assess multiple aspects of childhood outcomes over time, for the purposes of developing better policies and services for families. GUS contains two cohorts: the *Birth Cohort*, who have been tracked since they were babies (approximately 5,000 children at the start of data collection), and a *Child Cohort*, tracked from age 3 years (approximately 3,000 children at the start of data collection). The data here comes from the Birth Cohort, who have been tracked annually or bi-annually in testing sweeps up to the start of secondary school (i.e., Sweep 1 in 2005/06 occurred at age 10 months, with annual sweeps until the age of six, then bi-annual sweeps up to Sweep 9 which took place at the start of secondary school in 2017/18, when the children were 12-13 years). The data we present has been collected from children between the ages of 4 and 12 years; specifically across

Sweeps 5, 8 and 9, when the children were ~5 years, ~10 years, and ~12 years of age. In particular, children were screened for grapheme-colour synaesthesia in Sweep 9 and, once their status was known, they were compared retrospectively group-wise (synaesthetes vs. controls) in their earlier tests of language ability carried out between Sweeps 5 and 8.

In screening children for synaesthesia we will use the behavioural ‘gold standard’ test (e.g., Amin et al., 2011; Baron-Cohen, Wyke, & Binnie, 1987; Cytowic, 1995; Eagleman, Kagan, Nelson, Sagaram, & Sarma, 2007; Rich, Bradshaw, & Mattingley, 2005; Simner, Harrold, Creed, Monro, & Foulkes, 2009; Ward & Simner, 2003) which measures the consistency of colours over time (i.e., a defining quality of synaesthesia). In a typical test, researchers presents participants with a list of triggers (e.g., letters) and require participants to choose a colour for each trigger (e.g., A = red; B = blue). The test is then repeated some time later, without warning, and a comparison is made between the colours given during the test and retest. This determines whether the associations for each letter are consistent over time (e.g., A = red, both at test and retest). Synaesthetes are highly consistent (e.g., adults and children over 10-11 years are typically 70-100% consistent; Simner, 2012) despite the gap between test and retest (e.g., which can be anything from several minutes to many years; e.g., Simner & Logie, 2007). In contrast, average non-synaesthetes tend to be *inconsistent* in their colour choices. Our study uses this well-validated methodology to separate synaesthetes from non-synaesthetes, in combination with a second feature typically found in synaesthetes: that they can self-report having synaesthesia. We therefore tested children both in a consistency test, and with a self-diagnosing questionnaire asking whether participants have experienced long-term colours for letters and numbers. We use these measures to classify three types of children: *synaesthetes* score high in consistency *and* self-report synaesthesia in our questionnaire; *average memory non-synaesthete controls* do not score well in consistency and do not self-report synaesthesia;

and *high memory non-synaesthete controls* score high in consistency but do *not* claim to have synaesthesia (i.e., they are not synaesthetes but can remember relatively well the colours they gave previously for graphemes). These groupings are described more fully in our methods below, but by classifying children in this way we can separate the cognitive advantages of synaesthesia from those associated simply with having a good memory.

Finally, we also considered another feature from this dataset: the highest level of education of parents in the household. This is a known predictor of multiple child outcomes including academic achievement (e.g., Davis-Kean, 2005) so we will factor this into our model in order to better test outcomes from synaesthesia, independently of collateral features from parents. This measure was elicited from the GUS cohort in a parent questionnaire we name *Household Education Survey*, taken at multiple sweeps (we will use the Sweep immediately prior to that of our dependent measure; see *Methods*). In summary, we test whether synaesthetic children show superior abilities in expressive and receptive vocabulary, in sentence comprehension and in their academic self-concept (i.e., how they estimate their personal academic achievements), while factoring out superfluous influences such as parental education.

## **Methods**

### **Participants**

We tested 2037 children (51.6% female) from the GUS Birth Cohort 1. Our participant cohort was initially recruited by GUS from 130 geographical regions (known and *Primary Sampling Units* or PSUs) randomly selected from the geographical zones of Scotland. All households with children born between June 2004 and May 2005 within those PSUs were eligible for inclusion in the GUS birth cohort (one child per household). Testing has taken place over a

longitudinal testing period of 11 years, with our tests of interest carried out at Sweep 5 (expressive vocabulary; mean age 58.1 months; SD 0.4; henceforth “~5 years”), Sweep 8 (expressive/ receptive vocabulary and sentence comprehension; mean age 120.7 months; SD 2.6; henceforth “~10 years”) and Sweep 9 (grapheme-colour synaesthesia diagnostic; henceforth “~12 years”; see *Results* for demographic information on each group diagnosed<sup>1</sup>). The timeline of our testing is shown in Table 1. Attrition rates within the GUS study affected some demographic groups more than others, with fewer children from families with lower parental qualifications making up the Sweep 8 sample (5% compared to 9% of the Sweep 1 sample). Similarly the Sweep 8 sample was made up from a larger proportion of families with higher qualifications (36% compared to 28% of the Sweep 1 sample). Given this, we have taken care to control for parental education in our analyses. Our study had ethical clearance from Scotland ‘A’ MREC committee (Sweeps 1-8) and the NatCen Research Ethics Committee (Sweep 9), and from the local university ethics board at the University of Sussex.

### *Materials and Procedure*

At each sweep, children were tested in their homes by a GUS researcher in the presence of a parent or care-giver. As noted above, testing took place over 11 years and the screening for grapheme-colour synaesthesia was carried out immediately before the preparation of this paper (see ScotCen Social Research, 2018 for full details of data collection). Children completed the following tests of interest.

---

<sup>1</sup> Comparable age data from Sweep 9 has not yet been released by the GUS project. Please see a breakdown by group in the *Results*, and respective ages from Sweep 8 (i.e. the most recent GUS sweep with available data).

*Grapheme-colour synaesthesia test, (Sweep 9, age 12-13 years)*

Our test for grapheme-colour synaesthesia was presented during Sweep 9 of data collection for the GUS project (at age ~12years). This test (see Simner, Hughes, Carmichael, & Smees, 2019) comprises two presentations of materials, given approximately 20 minutes apart during a single testing session. At each presentation children were given an A4 sheet of paper showing 36 graphemes (A-Z, 0-9). Graphemes were placed one per line, in the left hand margin in a randomly ordered list (each child saw one of five possible randomisations). Alongside each grapheme was a row of 13 small colour patches (0.4 x 0.4 cm, against a grey background, RGB value 123 123 123). These 13 colours were the eight basic colour categories of English (red, blue, green, yellow, pink, purple, orange, brown) along with black, white, and grey. The order of these 13 colours was randomised from line to line. We used prototypical exemplars of each colour in RGB taken from Berlin and Kay (1969; with the exception of brown and grey which were adjusted marginally for print quality; see Simner et al., 2019). Participants were instructed to choose the ‘best’ colour for each letter or number by circling its colour patch. They were told there was no right or wrong answer but to avoid picking the same colour for everything. As an example, Figure 1 below shows the graphemes N, B, and 3 (with example answers).



**Figure 1.** Example of the grapheme-colour test and sample replies (the circled colours).

Children completed this colour-picker task twice, with a 20 minute gap separating the two presentations. During this gap, children completed other tasks for the GUS project, unrelated to our current interests. The second presentation of our colour-picker test was given at the end of the session, without prior warning. The colour-picker task was identical to before, but with

a new randomisation order per child, and we included an extra instruction as follows “*Here is an exercise you did earlier. We'd like you to do the exercise again. Remember there's no right or wrong answer, you can do it any way you want. We're asking you to do it twice because we need to capture two different moments in time.*” In our results we will inspect the consistency with which children gave colours for each grapheme across the two presentations, since high consistency is a marker of synaesthesia.

In addition to the colour-picker task, children were given a short self-report questionnaire, which comprised three questions (shown in Figure 2 below, along with the responses that would be typical of a synaesthete). The first two questions appeared at the end of the first presentation, while the third question appeared at the end of the second presentation. Given the colour-picker and question(s), each presentation took approximately 5-10 minutes to complete.

<p><b><u>Time 1</u></b></p> <p><b>Was this the FIRST TIME you've ever thought about the colours of letters and numbers? Or had you ALREADY thought about knowing their colours, even before today? (tick one box here)</b></p> <p><input type="checkbox"/> This is the first time I've thought about their colours</p> <p><input checked="" type="checkbox"/> I had already thought about knowing their colours, even before today</p> <p><b>Did you circle the same colours you'd thought about in the past? Or did you just decide the colours today? (tick one box here)</b></p> <p><input checked="" type="checkbox"/> These were the colours I'd thought about in the past</p> <p><input type="checkbox"/> I just decided these colours today</p> <p><b><u>Time 2</u></b></p> <p><b>Which sentence best describes you? (tick one box here)</b></p> <p><input checked="" type="checkbox"/> Since I was little, I've always thought letters and numbers had their own colours</p> <p><input type="checkbox"/> Since I was little, I've always thought letters had their own colours, but numbers don't</p> <p><input type="checkbox"/> Since I was little, I've always thought numbers had their own colours, but letters don't</p> <p><input type="checkbox"/> I've never thought letters and numbers had colours; I only thought about this today</p>
--

**Figure 2.** Self-report questionnaire for grapheme-colour synaesthesia. Questions at “Time 1” accompanied the first presentation of the colour-picker test, while the question at “Time 2” accompanied the second. Check-marks indicate responses befitting synaesthetes (see *Results*).

*Receptive Vocabulary, Expressive Vocabulary and Sentence comprehension; WIAT II UK Listening Comprehension (Sweep 8, age ~10 years)*

The *WIAT II UK Listening Comprehension* sub-test (Wechsler, 2005) was presented as part of Sweep 8 data collection of the GUS project (age ~10 years). This test was originally designed for ages 4-21 years and measures children's oral language skills in three separate tasks: *receptive vocabulary*, *expressive vocabulary* and *sentence comprehension*. The *receptive vocabulary* test consists of 16 items, increasing in difficulty. For each item the child is presented with four pictures and is asked to point to the picture that matches a word spoken by the examiner (e.g., "empty"). Testing is discontinued if the child gets three consecutive items incorrect. The *sentence comprehension* test follows a similar format; the child is presented 10 sentences, read aloud by the experimenter (e.g., "Grandma is walking upstairs to get her hat"). For each sentence the child is presented with four pictures and instructed to point to the picture described exactly by the sentence. Finally, the *expressive vocabulary* test includes 15 items, again progressively more difficult. For each item the child is shown one picture, and read a brief description of what is represented (e.g., children are shown a picture of an oasis, and are told "a fertile place in the middle of a desert"). Children are instructed to say one word that has the same meaning as the picture and description (here, the child says "oasis").

*Academic self-concept in Reading and Numeracy ability (Sweep 8; ~10 years)*

At Sweep 8 (age ~10 years) children were also asked two questions on academic self-concept within a longer questionnaire covering a broad range of topics related to academic and social well-being (see *Growing up in Scotland*; Scottish Executive., 2007). These two academic self-concept questions related to children's feelings about their abilities in reading and numeracy, and were phrased as: 'Are you good at reading?', and 'Are you good at number work?' Children

could give one of four possible responses: *Agree strongly*, *Agree*, *Disagree*, or *Disagree strongly*.

*Expressive Vocabulary (BAS II Naming Vocabulary) (Sweep 5, age ~5 years)*

The *BAS II Naming Vocabulary* test is a second test of expressive vocabulary, which was presented as part of Sweep 5 data collection of the GUS project (at ~5 years). This is a sub-test of the *British Ability Scales Second Edition* (BAS II; Elliott, Smith, & McCulloch, 1996) and is designed for use from 2 years 6 months into adulthood. The sub-test *Naming Vocabulary* assesses expressive vocabulary which is the task of interest in the current study. Children were presented with a series of up to 36 pictures of objects (e.g., a shoe), one at a time, and were asked to name them.

*Household Education Survey (Sweep 7 age ~8 years)*

Parents/Carers also completed a questionnaire in which they indicated their own level of education in response to the question “Have you passed any of the exams or got any of the qualifications on this card? [see Figure 3]”. Within single parent/carer households this question was completed for the sole respondent, but within couple-households it reflected the highest household education across both respondents. Parents were asked this question at multiple sweeps but we use Sweep 7 as the measure taken immediately prior to that of our primary measures (in Sweep 8). Any missing data from Sweep 7 was replaced by data from Sweep 8.



MaMedf02	Univ/CNAA first degree/diploma
MaMedf03	Postgraduate degree
MaMedf04	Teacher training qualification
MaMedf05	Nursing qualification
MaMedf06	Foundation/advanced modern apprenticeship
MaMedf07	Other recognised trade apprenticeship
MaMedf08	OCR/RSA - (Vocational) Certificate
MaMedf09	OCR/RSA - (First) Diploma
MaMedf10	OCR/RSA - Advanced Diploma
MaMedf11	OCR/RSA - Higher Diploma
MaMedf12	Other clerical, commercial qualification
MaMedf13	City&Guilds - Level 1/ Part I
MaMedf14	City&Guilds - Level 2/ Craft/ Intermediate/ Ordinary/ Part II
MaMedf15	City&Guilds - Level 3/ Advanced/ Final/ Part III
MaMedf16	City&Guilds - Level 4/ Full Technological/ Part IV
MaMedf17	SCOTVEC/BTEC First Certificate
MaMedf18	SCOTVEC/BTEC First/General Diploma
MaMedf19	SCOTVEC/BTEC/BEC/TEC (General/Ordinary) National Certif or Diploma (ONC/OND)
MaMedf20	SCOTVEC/BTEC/BEC/TEC Higher National Certif (HNC) or Diploma (HND)
MaMedf21	SVQ/NVQ Lev 1/ GSVQ/GNVQ Foundation lev
MaMedf22	SVQ/NVQ Lev 2/ GSVQ/GNVQ Intermediate lev
MaMedf23	SVQ/NVQ Lev 3/ GSVQ/GNVQ Advanced lev
MaMedf24	SVQ/NVQ Lev 4
MaMedf25	SVQ/NVQ Lev 5
MaMedf26	Professional qualification (employment related)
MaMedf27	IT certificate/qualification (other than those listed above)
MaMedf28	Other employment related qualification
MaMedf94	Other specific qualification

**Figure 3.** List of education or examination attainments within the *Household Education Survey*. Left column shows the GUS classification number for each level of award.

Within our results, these qualifications are converted into the following six categories (following GUS; Anderson et al., 2007) based on the *Scottish Credit and Qualifications Framework* (please see “Scottish Credit and Qualifications Framework,” 2006): 1= 'no qualification'; 2 = 'other'; 3 = 'Lower level Standard Grades and Vocational qualifications'; 4 = 'Upper level Standard Grades and Intermediate Vocational qualifications'; 5 = 'Higher grades and Upper level vocational qualifications'; 6 = 'Degree level academic and vocational qualifications'.

Tables 1 below shows the timeline for data collection from sweeps 5-9. Data collection for this GUS birth cohort began when children were 10 months old (Sweep 1), but this study focusses on information from Sweep 5 and higher.

GUS sweep Year data collected	Child's age	Test
<b>Sweep 5</b> 2009/10	4-5 years	Expressive Vocabulary test

<b>Sweep 6</b> 2010/11	5-6 years	
<b>Sweep 7</b> 2012/13	7-8 years	Household Education Survey
<b>Sweep 8</b> 2014/15	10-11 years	Expressive Vocabulary test Receptive Vocabulary test Sentence comprehension test Academic self-concept self-report (Maths/Reading)
<b>Sweep 9</b> 2017/18	12-13 years	Synaesthesia Screening

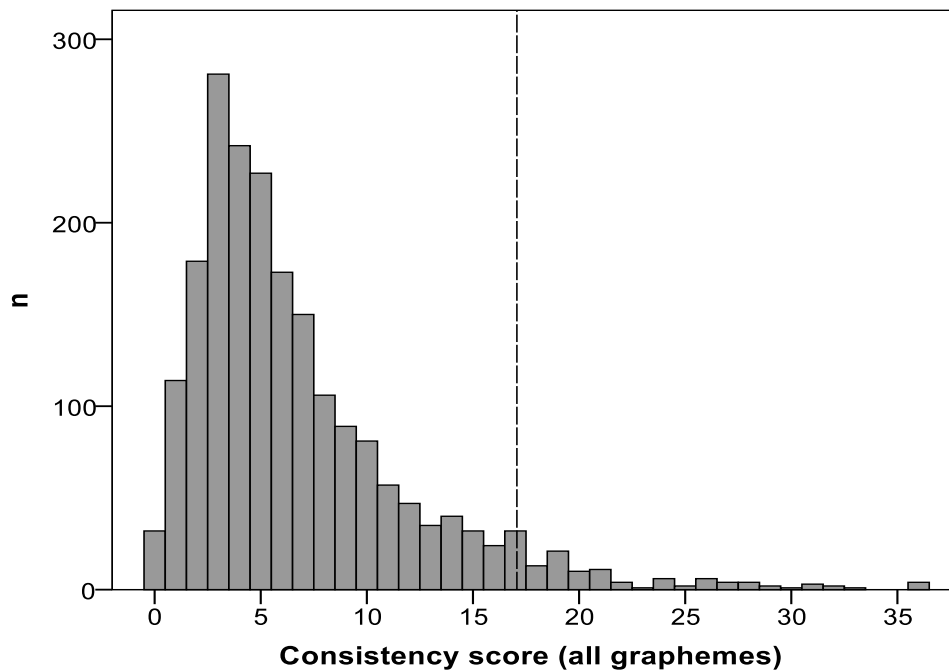
**Table 1.** Timeline of GUS data collection for data used in present study.

## *Results*

### *Identifying Synaesthetes*

Synaesthesia identification was based on two pieces of information: performance in the self-report questionnaire and performance in the objective test of consistency, both collected at Sweep 9. In self-report, children who answered positively to at least two of the questions (i.e., stated they had synaesthesia and confirmed at least once more) were considered ‘self-report’ synaesthetes (but this was not yet a definitive diagnosis). Second, we calculated the number of graphemes where the child picked the same colour at both presentation 1 and 2. When a letter or number was given the same colour twice (e.g., the number 7 was given blue on both occasions) we scored this 1 point, and calculated a total consistency score for letters (out of 26), for numbers (out of 10) and for letters and numbers (out of 36). Children passed the consistency test if their scores were 1.96 standard deviations or more above the mean (i.e. significantly higher than other children in the sample). This 1.96 cut-off is in line with methodology used in previous research in children (Simner et al., 2009) and corresponds to being 50% consistent for letters, 70% consistent for numbers; and 50% consistent across all 36 graphemes. Figure 4 shows the distribution of scores in consistency for graphemes (letters and numbers combined). To contextualise this threshold, the probability of getting 48% consistent graphemes by chance is  $< 0.000001$ , and random responding would have produced only 6

consistent responses. Finally, we point out that studies using more complex colour pickers (Carmichael, Down, Shillcock, Eagleman, & Simner, 2015; Rothen, Seth, Witzel, & Ward, 2013) are able to establish their cut-off using distances in colour space, but we could not rely on this method here (i.e., no fine-grained colour palette).



**Figure 4** Distribution of consistency scores for all graphemes (i.e., numbers and letters, maximum possible score 36). The dotted line shows the threshold above which a synaesthete is recognised.

To be classified as *synaesthetes*, children had to pass at least one of the grapheme-colour consistency tests (i.e., score higher than the cut-offs for number only, or letter only, or all 36 graphemes) and pass the self-report questionnaire (i.e., report synaesthesia in at least two of the three self-report questions). Average memory non-synaesthete controls (henceforth *Average memory controls*) were children who failed all consistency tests, and answered negatively to at least two of the three questions answered. Finally, high memory non-synaesthete controls (henceforth *High memory controls*) were a small group of children who passed at least one of the consistency tests but answered negatively to at least two of the three

questions (i.e., reported they did not have synaesthesia on the majority of questions but gave consistent colours, likely from good memory alone; see Simner et al., 2009). We point out that we categorised children on two out of three of the self-report questions (e.g., two out of three confirmations for a synaesthetes; or two out of three rejections for a non-synaesthete) in order to allow children a small amount of inattention when completing our task. However, we point out that the pattern of results shown below remains the same whether we categorise children according to two questions or three questions.

Following our screening for synaesthesia, our final sample used in the analyse below comprised 51 grapheme-colour synaesthetes<sup>2</sup> (51% female; mean age Sweep 8<sup>3</sup> = 121.1 months, range 116-126, *S.D.* = 2.5), 1407 Average memory controls (50% female; mean age Sweep 8 = 120.7, range 114-127, *S.D.* = 2.6) and an additional 73 High memory controls (52% female; mean age Sweep 8 = 121.1, range 114-127, *S.D.* = 2.5). There were no gender difference between the three groups,  $\chi^2(2, N = 1531) = 0.099, p = .95$ . In addition there were no age differences between groups at Sweep 8 (when our primary dependent variables were measured),  $f(2, 1531) = 1.539, p = .22$ .

An additional 506 children were excluded from our study. Of these, 153 did not have enough literacy or demographic data from Sweeps 7 or 8 to be included in our analyses. Another 344 children were removed because their synaesthesia status was ambiguous in that they failed the synaesthesia consistency test but self-reported synaesthesia at least twice. (In contrasts, our

---

<sup>2</sup> For clarity we point out that all 51 children had letter-colour synaesthesia and/or number-colour synaesthesia. We also identified three additional synaesthetes who had achieved their status across 36 consistent graphemes (while failing consistency out of 26, and out of 10). However, these synaesthetes were part of the cohort described below, whom we removed for missing historical data.

<sup>3</sup> Age at sweep 9 was not available at the time of analysis.

synaesthetes acted and self-reported like synaesthetes, while our controls acted and self-reported like high/average controls). This phenomenon of false reporting of synaesthesia has been found elsewhere with adults (Simner, Ipsier, Smees, & Alvarez, 2017), and the complex possible reasons for this have been discussed in detail by Simner (2019). Lastly, the remaining 9 children were removed because they failed to follow task instructions in that they chose the same colour or achromatic (grey, black, white) for at least 50% of their consistent graphemes.

### *Language Abilities of Synaesthetes and Non-synaesthetes*

In the following three sections we consider the abilities of synaesthetic children and controls in three ways: in mid-childhood (~10 years), in early childhood (~5 years), and from the perspective of improvement across these time intervals. In the BAS *Naming Vocabulary* test we used standardised scores that were provided within the measure; otherwise we used raw scores while controlling for age within our models.

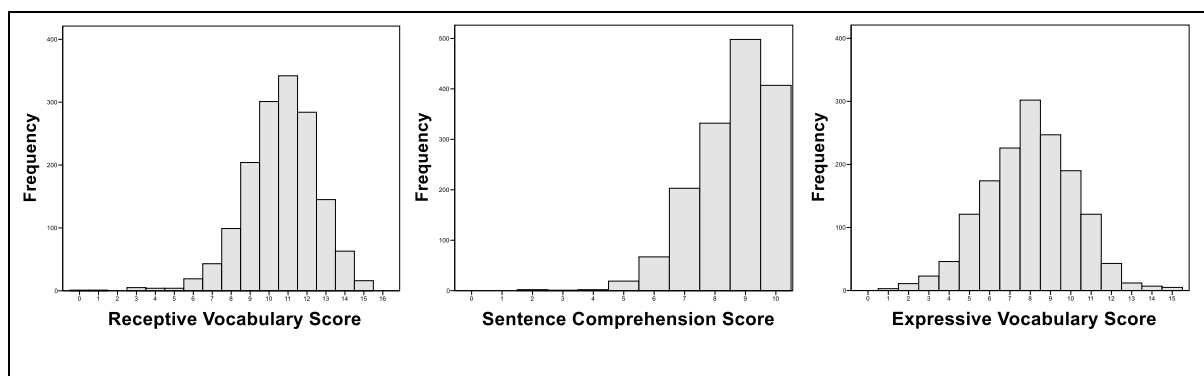
#### *Mid-childhood language ability age ~10 years: Vocabulary, sentence Comprehension, and academic self-concept*

We considered children's scores in three formal tests of language development as well as their own ratings of academic self-concept, all taken in Sweep 8 (i.e., one year prior to their synaesthesia screening). The three formal tests were of *receptive vocabulary*, *sentence comprehension*, and *expressive vocabulary* taken from the WIAT II UK Listening Comprehension subtests (Wechsler, 2005)<sup>4</sup>. Descriptive statistics and the overall distribution

---

<sup>4</sup> There was only weak to modest association between the three sub-tests suggesting, as expected, they measure different domains of language development. Receptive and expressive vocabulary showed the strongest association ( $r = .37, p < .01$ ), followed by expressive vocabulary and sentence comprehension ( $r_s = .31, p < .01$ ). The weakest correlation was between receptive vocabulary and sentence comprehension ( $r_s = .22, p < .01$ ).

of data for each test is shown in Table 2 and Figure 5 below. Data from the receptive vocabulary test (age ~10 years), and sentence comprehension (age ~10 years) failed the Kilmorgov-Smirnoff test of normality (normality of residuals). As receptive vocabulary was only moderately skewed, parametric analyses were carried out, with additional bootstrapping. However, sentence comprehension was considered too highly skewed for parametric analyses so instead a binary dependent variable was derived. We derived this binary variable by taking children who scored on the 90<sup>th</sup> percentile or above (i.e. top ten percent of the sample) versus the rest of the sample.



**Figure 5.** Distribution of language development measures at age ~10 years

	Mean (SD)	Median (inter-quartile range)	Range
Receptive Vocabulary	10.63 (1.90)	11 (3)	0-15
Sentence Comprehension	8.58 (1.25)	9 (2)	0-10
Expressive Vocabulary	7.99 (2.21)	8 (2)	0-15

**Table 2.** Descriptive statistics for language development measures at age ~10 years

We analysed our data primarily with mixed effects linear or binary logistic regression, with alpha set at  $p < .05$  and effect sizes shown as Cohen's  $d$  ( $d$ )<sup>5</sup> or Odds Ratios (OR). All analyses were computed in SPSS 24.0. For both expressive and receptive vocabulary (which we analyse with Mixed Linear effects models) and for sentence comprehension and academic self-concept (which we analyse with Mixed Binary Logistic effects models), we cluster children with PSUs.

<sup>5</sup> The Cohen's  $d$  formula is adapted for multilevel analyses, by Elliot and Sammons (Elliot & Sammons, 2004). Henceforth throughout the paper referred to as Cohen's  $d$ .

As a reminder, PSUs, or *Primary Sampling Units*, are the 130 geographical zones from which children were recruited. Although modelling this type of naturally-occurring clustering is more important when cluster-level itself is being investigated, or is high, we take this as a precautionary step since the impact of geographic clustering on development of synaesthesia is unknown. Where non-significant cluster variation was found, single level regression models were used. Age (in months) and gender were controlled for as covariates where significant (see below). For example, listening comprehension subtests have no available age standardisation. Since age was found to be significantly related to both receptive and expressive vocabulary in our sample ( $r = .10$  and  $r = .12$  respectively), we included age in the statistical model as a covariate. Likewise, we also included as co-variate the highest household qualification where significant (i.e., whenever household qualification significantly predicted the dependent variable; this was the case for all three vocabulary and language skills outcomes at age 10). The highest of the six levels of education (i.e., “Degree level ...” see above) was treated as the comparison group for the remaining five, which were included in the model as dummy variables<sup>6</sup>. Finally, in all models within the main text we treat our largest group (*average memory controls*) as our comparison group (i.e., we compared *average memory controls* separately to *synaesthetes* and to *High memory controls*). The SI contains our additional models where *synaesthetes* and *high memory controls* are compared directly.

---

<sup>6</sup> Our statistics are robust to differences in group size (between controls, high memory and synaesthetes) because sample size within groups is adequate (Wilson Vanvoorhis & Morgan, 2007) and all pass test for homogeneity of variance between groups (Field, 2009). Cognitive attainment is known to show strong clustering effects at the school or organisation level (Creemers & Kyriakides, 2010; Sammons et al., 2015, 2008) and also at the area level (Bramley & Kofi Karley, 2007; Garner & Raudenbush, 1991; Melhuish, 2010; Raviselvam, Anderson, Holtta-Otto, & Wood, 2018). We therefore choose to model the natural clustering of the data within a multilevel framework, utilising mixed hierarchical linear or logistic models. As a randomly sampled dataset we were keen to retain as many children as possible to keep it as representative as possible, so chose not to utilise a matching procedure, typical of RCTs or lab-based studies. All dependent continuous variables passed homogeneity of variance tests.

Tables 3-5 shows our regression models for language scores at age ~10 years. We found that *synaesthetes* outperformed *high memory controls* for receptive vocabulary ( $d = 0.34$ , bootstrapped  $p < .05$ ) and expressive vocabulary ( $d = 0.47$ ,  $p < .001$ ) but there was no difference between *high memory controls* and *synaesthetes* in sentence comprehension ( $OR = 0.98$ ,  $p < .10$ ), even after controlling for household qualifications.

		<i>Receptive Vocabulary</i>	
		Estimate (SE)	d (95% CI)
Intercept		10.77 (0.22)**	
Age~		0.06 (0.02)**	0.19 (0.08,0.28)
Gender: Boys (vs Girls)		0.75 (0.09)**	0.43 (0.32,0.52)
Highest household qualification (vs degree)	None	-1.76 (0.29)**	-1.01 (-1.38,-0.59)
	Other	-0.13 (0.73)	-0.08 (-0.63,0.63)
	Lower	-0.66 (0.29)*	-0.38 (-0.73,-0.04)
	Upper	-0.80 (0.14)**	-0.46 (-0.62,-0.28)
	Higher	-0.21 (0.11)*	-0.12 (-0.23,-0.01)
Synaesthete (vs High memory)		0.60 (0.33)*	0.34 (0.01,0.66)
Average Memory (vs High memory)		-0.30 (0.21)	-0.17 (-0.36, 0.06)
Random: PSU		0.21 (0.06)***	
Random: Child		3.05	
N		(0.12)***	
		1531	

**Table 3:** Mixed Linear effects models for Receptive Vocabulary at Sweep 8 (parentheses show standard error). Synaesthetes and average memory controls are compared to high memory controls. \* $p < 0.05$  \*\* $p < 0.01$  \*\*\*  $p < 0.001$ , ~grand centred mean, nf not fitted. d Cohen's d; Bootstrapped significance values shown

		<i>Sentence Comprehension</i>	
		B (SE)*	OR (95% CI)
Intercept		-0.41 (0.27)	
Age~		nf	
Gender: Boys (vs Girls)		-0.38 (0.12)**	0.66 (0.53,0.86)
Highest household qualification (vs degree)	None	-0.72 (0.46)	0.49 (0.20,1.21)
	Other	0.09 (0.90)	1.09 (0.19,6.37)
	Lower	-0.67 (0.46)	0.51 (0.21,1.27)
	Upper	-0.19 (0.18)	0.83 (0.58,1.18)
	Higher	-0.35 (0.13)*	0.71 (0.54,0.93)
Synaesthete (vs High memory)		-0.02 (0.40)	0.98 (0.45, 2.16)
Average Memory (vs High memory)		-0.29 (0.27)	0.74 (0.44,1.25)
Random: PSU		0.25 (0.09)**	
Random: Child		---	
N		1531	

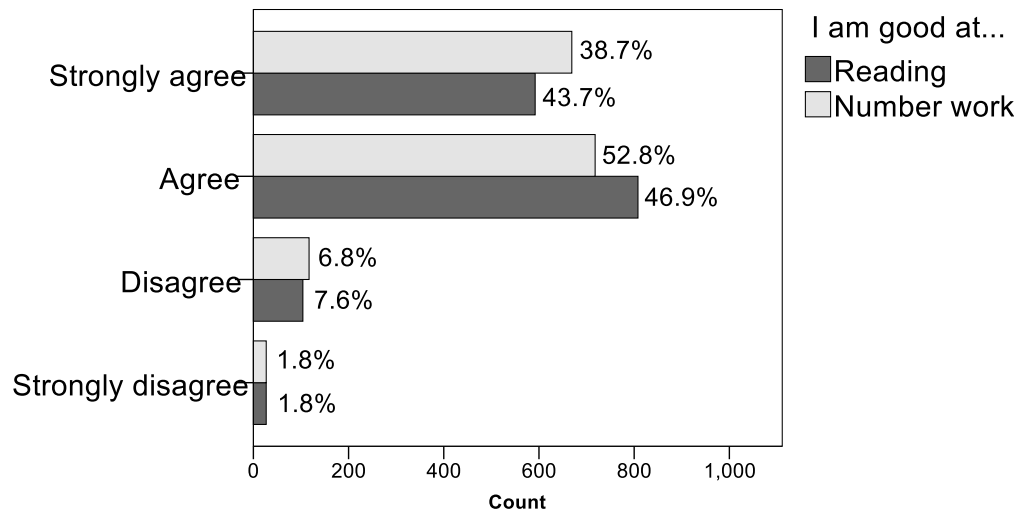


**Table 4:** Mixed Binomial logistic models for Sentence comprehension at Sweep 8 (parentheses show standard error). Synaesthetes and average memory controls are compared to high memory controls. \* $p < 0.05$  \*\* $p < 0.01$  \*\*\*  $p < 0.001$ , ~grand centred mean, nf not fitted. d Cohen's d

		<i>Expressive Vocabulary</i>	
		Estimate (SE)	d (95% CI)
Intercept		8.61 (0.26)***	
Age~		0.10 (0.02)***	0.26 (0.15,0.35)
Gender: Boys (vs Girls)		nf	
Highest household qualification (vs degree)	None	-1.72 (0.35)***	-0.84 (-1.17,-0.51)
	Other	-1.27 (0.86)	-0.62 (-1.44,0.20)
	Lower	-0.81 (0.35)*	-0.40 (-0.73,-0.06)
	Upper	-0.79 (0.16)***	-0.38 (-0.54,-0.23)
	Higher	-0.51 (0.12)***	-0.25 (-0.37,-0.13)
Synaesthete (vs High memory)		0.96 (0.38)*	0.47 (0.10,0.84)
Average Memory (vs High memory)		-0.32 (0.25)	-0.16 (-0.40,0.08)
Random: PSU		0.37 (0.10)***	
Random: Child		4.20 (0.16)***	
N		1531	

**Table 5:** Mixed Linear effects models for Expressive vocabulary at Sweep 8 (parentheses show standard error). Synaesthetes and average memory controls are compared to high memory controls. \* $p < 0.05$  \*\* $p < 0.01$  \*\*\*  $p < 0.001$ , ~grand centred mean, nf not fitted. d Cohen's d

We next considered responses from children's own academic self-concept at age ~10 years, shown descriptively in Figure 6 and Table 6 below. As the data was highly skewed the dependent variable was converted into a binary response: Strongly agree (=1) versus all other responses (= 0). *Synaesthetes* had a significantly higher academic self-concept compared to *high memory controls* for reading (OR = 2.34,  $p < .05$ ) but not in their self-concept for numeracy (OR = 0.89,  $p < .10$ ), even after controlling for household qualifications.



**Figure 6.** Children's Academic self-concept responses at age ~10 years

		<i>Academic Self-Concept: Reading</i>		<i>Academic Self-Concept: Number work</i>	
		Estimate (SE)	OR (95% CI)	Estimate (SE)	OR (95% CI)
Intercept		-0.30 (0.25)		-0.62 (0.25)***	
Age~		nf		nf	
Gender: Boys (vs Girls)		nf		0.75 (0.10)***	2.12(1.73,2.60)
Highest household qualification (vs degree)	None	-0.96 (0.39)*	0.38(0.18,0.82)	nf	
	Other	0.11 (0.83)	1.12(0.22,5.69)	nf	
	Lower	-0.85 (0.38)*	0.43(0.21,0.90)	nf	
	Upper	-0.75 (0.17)***	0.47(0.34,0.66)	nf	
	Higher	-0.31 (0.12)**	0.74(0.58,0.93)	nf	
Synaesthete (vs High memory)		0.85 (0.38)*	2.34(1.12,4.90)	-0.12 (0.38)	0.89(0.42,1.85)
Average Memory (vs High memory)		0.06 (0.25)	1.06(0.65,1.73)	-0.02 (0.25)	0.98(0.61,1.59)
Random: PSU		nf		nf	
N		1531		1531	

**Table 6:** Binary logistic models for Academic self-concept in Reading and Number Work at Sweep 8 (parentheses show standard error) Synaesthetes and average memory controls are compared to high memory controls. \* $p < 0.05$  \*\* $p < 0.01$  \*\*\*  $p < 0.001$ , ~grand centred mean, nf not fitted.

### *Does the sub-type of synaesthesia predict language skills?*

To investigate this question, post-hoc analyses were carried out on measures where significant group-wise differences had been found. We first categorised synaesthetes according to their synaesthetic trigger into three mutually exclusive groups, i.e., as: *number-only synaesthetes* (n=12), *letter-only synaesthetes* (n=20) and *both letter and number synaesthetes* (i.e., children

who had both types of synaesthesia at once;  $n=19$ ). The *average memory controls* were also included in the analysis as a fourth separate category, and all four were compared to the *high memory control* group. Age at test (listening comprehension sub-tests, Sweep 8), gender and household qualifications were again controlled for where appropriate (see SI for full statistical models).

We found that children with *number-only synaesthesia* performed significantly better than high memory controls in both vocabulary measures (Receptive  $d = 0.58$ ,  $p < .01$ ; Expressive vocabulary  $d = 0.88$ ,  $p < .01$ ). Children with *letter-only synaesthesia* did not perform significantly better than high memory controls in either vocabulary measures (Receptive  $d = 0.11$ ,  $p > .05$ ; Expressive vocabulary  $d = 0.21$ ,  $p > .05$ ) although it is likely that the sample size had some impact on our analysis of smaller sub-groups ( $n=12$ ; 21; 19 respectively for *number-only synaesthetes*, *letter-only synaesthetes* and *both number- and letter-colour synaesthetes*). For academic concept for Reading Academic self-concept, it was *letter-only synaesthetes* who reported significantly better academic self-concept than average memory controls (OR = 2.83,  $p < .05$ ).

### *Earlier language development*

Expressive vocabulary (i.e., *BAS II Picture naming*) was tested not only at age ~10 years, but also at age ~5 years. We therefore looked at whether synaesthetes showed advantages over their peers even at this earlier stages in development (based on  $n=1486$ , given that some children did not complete this test age ~5 years). In addition, we investigated their *progress* in language skills by considering their rate of improvement over time (i.e., comparing expressive

vocabulary collected at ~5 years against ~10 years, which represents the beginning and end of primary school).

Our 1486 children were 51% female (mean age of testing at start of primary school = 58.1, range 57-60, S.D.= 0.4; mean age of testing at end of primary school = 120.7, range 114-127, S.D.= 2.6). Split by synaesthesia classification, the analysis included 50 *grapheme-colour synaesthetes* (52% female; mean age Sweep 8 = 121.0, range 116-126, S.D. = 2.4), 1367 *average memory controls* (50% female; mean age Sweep 8= 120.7, range 114-127, S.D. = 2.6) and 69 *high memory controls* (52% female; mean age Sweep 8 = 121.1, range 114-127, S.D. = 2.6). There was again no age difference between the three groups ( $f(2, 1486) = 1.12, p = .32$ , or gender,  $\chi^2 = (2, N = 1486) = 0.174, p = .92$ ).

We incorporated expressive vocabulary skills, measured by BAS II *Naming Vocabulary* subtest into our earlier mixed effects models. Table 7 below shows the model for progress in expressive vocabulary across early to mid-childhood period (age ~5 to age ~10 years). As expected, age and parental education were significant predictors of vocabulary (Davis-Kean, 2005), with older children and children from more highly education households making more progress over time. Importantly, we also found that *synaesthetes* made more progress than *high memory controls* (see Table 7) even after controlling for age and household qualifications.

		<i>Progress Sweep 5 to Sweep 8</i>	
		Coefficient (SE)	d
Intercept		8.42 (0.24)***	
Age~		0.11 (0.02)***	0.30 (0.19,0.39)
Gender: Boys (vs Girls)		nf	
Highest household qualification (vs degree)	None	-0.86 (0.34)*	-0.45 (-0.80,-0.10)
	Other	-1.14 (0.80)	-0.60 (-1.42, 0.22)
	Lower	-0.11 (0.34)	-0.06 (-0.40,0.29)
	Upper	-0.46 (0.15)**	-0.24 (-0.40,-0.08)
	Higher	-0.30 (0.12)*	-0.16 (-0.28,-0.04)

Expressive vocab, age 4-5~		0.08 (0.01)***	0.83 (0.75,0.96)
Synaesthete (vs high memory)		0.75 (0.36)*	0.39 (0.02,0.77)
Average memory (vs high memory)		-0.26 (0.24)	-0.13 (-0.38, 0.11)
Random: PSU		0.30 (0.08)***	
Random: Child		3.63 (0.14)***	
n		1486	

**Table 11.** Mixed effects Linear regression model for Expressive vocabulary: progress between Sweeps 5 and 8 (parentheses show standard error). \* $p < 0.05$  \*\* $p < 0.01$  \*\*\*  $p < 0.001$ , ~grand centred mean. Synaesthetes and average memory controls are compared to high memory controls.  $d$  Cohen's  $d$  (size of effect only)

Given this difference in progress we also checked whether synaesthetes *began* primary school with a pre-existing advantage by looking at their test scores age ~5 years in isolation. This analysis is presented in our SI but does not show any significant *a priori* advantage for synaesthetes at age ~5 years compared to *high memory controls* ( $d = 0.06$ ,  $p > .10$ ).

## Discussion

Here we have asked whether children with synaesthesia show enhanced cognitive linguistic skills compared to their peers. We screened a large cohort of more than two thousand children for grapheme-colour synaesthesia and identified a sample of more than 50 synaesthetes to compare with two types of non-synaesthete controls. *Average memory controls* were non-synaesthetes who performed at average levels when recalling coloured letters and numbers, while *high memory controls* were children who did not have synaesthesia but still performed well when recalling them (i.e., they gave consistent colours for graphemes from memory alone). Next, we then examined three retrospective tests of language development which had been measured prior to our screening (expressive vocabulary, receptive vocabulary, sentence comprehension), as well as two domains of academic self-concept measured at the same time (for reading and numeracy). We found modest but significant advantages for synaesthetes in both vocabulary and academic self-concept for reading. Specifically, *synaesthetes* performed

better than both *average memory* and *high memory controls* in expressive vocabulary, receptive vocabulary and in academic self-concept for reading. There was no significant advantage for synaesthetes in sentence comprehension nor in self-concept for number skills. Although effect sizes were positive, *high memory controls* did not perform statistically better than *average memory controls* on any domain, suggesting their memory skills were distinct from their linguistic skills.

Previous research in adults had suggested that synaesthesia can endow enhanced cognitive abilities and our own research supports these findings in children. Adult effect sizes (Cohen's *d*) are often only within the moderate range (Chun & Hupé, 2016; Gross et al., 2011; Rothen & Meier, 2010; Rothen et al., 2012; Yaro & Ward, 2007; see Rothen et al 2012 for review) and likewise, our own effects were significant but not exceptional (e.g., moderate effect sizes in both receptive and expressive vocabulary) also mirroring those found in synaesthetic children elsewhere for processing speed (Simner & Bain, 2017; see SI). Together our findings suggest, in line with Rothen and Meier (2010) for adults, that child synaesthetes may show modest advantages in multiple domains, but exceptional skills in fewer.

The advantages we found for synaesthetes in receptive vocabulary at age ~10 years are also largely in line with previous vocabulary findings by Green and Goswami (2008) in synaesthetes aged 7-15 years, and Simner and Bain (2018) in synaesthetes aged 10-11 years. But whereas one study recruited a self-referred sample (who might have performed well for reasons other than synaesthesia) the other study had a very small sample size and only a non-significant trend. Our own sample here had increased statistical power to identify even moderate effect sizes and had no self-selecting bias for synaesthetes over controls. This is not to say that our sample has

no self-selection bias whatsoever (after all, parents had to sign up to join the GUS birth cohort we studied), but importantly, both synaesthetes and controls came from the same sample so would be subject to the same recruitment influences. Despite this, synaesthetes still showed advantages over their non-synaesthetic peers. It might yet be argued that what we have found represents a type of interaction between synaesthetic status and motivation, in that advantages for synaesthetes over non-synaesthetes could emerge only when both types of children come from motivated families (i.e., families who enter research cohorts). We acknowledge this unusual but theoretically possible interaction and for this reason are also screening children for synaesthesia within a cohort of >3000 children who are entirely randomly recruited in that their parents needed to do nothing whatsoever to enter our study. That novel study is in progress and will represent an interesting test of the more fine-grained hypotheses we have raised here.

We also investigated developmental trajectories by comparing expressive vocabulary from the same children at two different ages: ~5 years (roughly in line with the start of primary schooling) up to the age of ~10 years (roughly in line with the end of primary schooling). We found synaesthetes made more progress across primary schooling than both average memory controls and high memory controls (i.e., improved at a steeper rate). Importantly, synaesthetes did not show an a priori advantage over average or high memory non-synaesthetes at age ~5 years. Our study is the first to assess how learning trajectories differ for synaesthetes and non-synaesthetes, and suggests that the development of synaesthesia goes hand in hand with the development of other (here, language) skills. Vocabulary acquisition is ongoing across the lifespan (Kuhl, 2010), but has its fastest acquisition prior to school entry (Nazzi & Bertoncini, 2003). It is therefore reasonable to expect that any differences which emerge during primary schooling (from age ~5 years), would continue into adulthood. This has been suggested by Chun and Hupé (2017), who found similar advantages in vocabulary knowledge for adult

synaesthetes (with an effect size of 0.53 but with statistical uncertainty from large confidence intervals; CI = 0.0, 1.0). Our results provide support for their important findings and suggest that advantages even stem back into childhood. Beyond Chun and Hupé, we also show these effects can be tied to grapheme-colour synaesthetes in particular, and that they extend across both receptive and expressive vocabulary knowledge separately.

Given our findings we might ask *why* synaesthetes are showing advantages in these domains. As noted in the introduction, one possible explanation for the literacy advantages of synaesthetes may rest on a *Dual-coding hypothesis* (Clark & Paivio, 1991; Paivio, 1969). In this account, synaesthetes may have improved memory (e.g., for vocabulary items) because these are encoded with additional information (here, colours) making their memory representations more robust. Evidence for this theory has come from studies showing, for example, that a group of letter-colour synaesthetes had superior letter-spans, but not number-spans (Radavansky et al 2011). Here we directly tested a Dual-coding account of our findings by assessing whether verbal advantages (i.e., linked to letters) might be found especially for children with coloured letters but not numbers. Our data did not support this hypothesis: post-hoc tests showed that superior performers were not only synaesthetes with coloured letters and numbers, but even those with coloured numbers only. This suggests that advantages go beyond improved memory from dual-coding alone. Likewise, we also found that the benefits for synaesthetes could not be accounted for by memory in a second way: synaesthetes outperformed age matched children not only with *average* memory spans (as measured by their consistency in pairing colours with graphemes) but also *above-average* memory. This suggests that the locus of synaesthetic advantages does not rest on this type of memory difference alone.



In place of Dual-Coding, what other mechanisms might account for our results? And how might skills and benefits interact with synaesthesia via genetics or the environment? Firstly, we know that synaesthesia is likely to be genetically primed (Asher et al., 2009; Rich, Bradshaw, & Mattingley, 2005; Tilot et al., 2018; Tomson et al., 2011; Ward & Simner, 2005) but whether it appears, and what form it takes, may be dependent on early experiences (Newell & Mitchell, 2016; Watson et al., 2017). Once it emerges, synaesthesia appears to correlate with other cognitive differences such as those identified here. Importantly, our study found that although differences in vocabulary ability were apparent at age 10, they were non-significant at age 5 years, a period when grapheme-colour synaesthesia is thought to still be developing (Simner et al., 2009). This suggests that vocabulary advantages may develop hand-in-hand with synaesthesia. However, the *direction* of any link between synaesthesia and cognition is difficult to deduce, and there are a number of ways to interpret our results.

Firstly, it may be that synaesthetic children, as with adults, have a preference for verbal information over visual information (Meier & Rothen, 2013). If so, this itself could aid the development of linguistic skills across the early school year, potentially boosting intellectual development in areas such as verbal ability (Rouw & Scholte, 2016). Alternatively, the global differences seen in the brains of synaesthetes (see Rouw, Scholte, & Colizoli, 2011) could give rise to skills such as an enhanced ability to organise or retrieve cognitive information more generally (Ramachandran & Azoulay, 2006). This could facilitate (inter alia) the storage and retrieval of words. Finally, the relationships between synaesthesia and cognition might instead be in reverse. Watson, Akins, Spiker, Crawford and Enns (2014) have suggested that synaesthesia may emerge in response to some type of early learning challenge. If they are correct, this ‘challenge’ could include new linguistic learning (e.g., reading and writing) at the start of schooling. If so, this could trigger synaesthesia (in those genetically predisposed) and perhaps especially in those who have had the greatest engagement in their learning. Future

work on other samples could look within the dataset for potential early learning challenges that might explain the development of grapheme-colour synaesthesia.

In conclusion we have shown that children with grapheme-colour synaesthesia performed better in tests of vocabulary, but not sentence comprehension, compared to their peers. Importantly, this effect arises even when other factors like age, recruitment and parental education levels are taken into account. Child synaesthetes also report higher beliefs of achievement in reading, but not in their number use, which may in part reflect a bidirectional relationship between skills and self-concept, each bolstering the other. Finally, our data suggest that children with synaesthesia do not start primary school *a priori* more advanced than their peers but they advance more quickly. In sum, these data suggest overall benefits for children with synaesthesia, at the level of lexical processing and reading.

### *References*

- Amin, M., Olu-Lafe, O., Claessen, L. E., Sobczak-Edmans, M., Ward, J., Williams, A. L., & Sagiv, N. (2011). Understanding grapheme personification: A social synaesthesia? *Journal of Neuropsychology*, *5*(2), 255–282. <https://doi.org/10.1111/j.1748-6653.2011.02016.x>
- Anderson, S., Bradshaw, P., Cunningham-Burley, S., Hayes, F., Jamieson, L., MacGregor, A., ... Wasoff, F. (2007). *Growing Up in Scotland: first research report on Sweep 1*. Edinburgh: Scottish Executive.
- Baron-Cohen, S., Wyke, M. A., & Binnie, C. (1987). Hearing words and seeing colors - an experimental investigation of a case of synesthesia. *Perception*, *16*(6), 761–767. <https://doi.org/10.1068/p160761>

- Berlin, B., & Kay, P. (1969). *Basic color terms : their universality and evolution*. University of California Press.
- Bramley, G., & Kofi Karley, N. (2007). Housing Studies Homeownership, Poverty and Educational Achievement: School Effects as Neighbourhood Effects.  
<https://doi.org/10.1080/02673030701474644>
- Brang, D., & Ramachandran, V. S. (2008). Psychopharmacology of synesthesia; the role of serotonin S2a receptor activation. *Med Hypotheses*, 70(4), 903–904.  
<https://doi.org/10.1016/j.mehy.2007.09.007>
- Carmichael, D. A., Down, M. P., Shillcock, R. C., Eagleman, D. M., & Simner, J. (2015). Validating a standardised test battery for synesthesia: Does the Synesthesia Battery reliably detect synesthesia? *Consciousness and Cognition*, 33, 375–385.  
<https://doi.org/http://dx.doi.org/10.1016/j.concog.2015.02.001>
- Chun, C. A., & Hupé, J.-M. (2016). Are synesthetes exceptional beyond their synesthetic associations? A systematic comparison of creativity, personality, cognition, and mental imagery in synesthetes and controls. *British Journal of Psychology (London, England : 1953)*, 107(3), 397–418. <https://doi.org/10.1111/bjop.12146>
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210. <https://doi.org/10.1007/BF01320076>
- Creemers, B., & Kyriakides, L. (2010). School Factors Explaining Achievement on Cognitive and Affective Outcomes: Establishing a Dynamic Model of Educational Effectiveness. *Scandinavian Journal of Educational Research*, 54(3), 263–294.  
<https://doi.org/10.1080/00313831003764529>
- Cytowic, R. (1995). *Synesthesia : Phenomenology And Neuropsychology A Review of*

Current Knowledge. *Psyche*, 2(10), 2–10.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment.

*Journal of Family Psychology*, 19(2), 294–304. <https://doi.org/10.1037/0893-3200.19.2.294>

Dunn, L. ., Dunn, L. ., Whetton, C., & Pintilie, D. (1982). The British Picture Vocabulary Scale : manual for the short and long forms. Windsor : NFER-Nelson.

Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D., & Sarma, A. K. (2007). A standardized test battery for the study of synesthesia. *Journal of Neuroscience Methods*, 159(1), 139–145. <https://doi.org/10.1016/j.jneumeth.2006.07.012>

Elliot, K., & Sammons, P. (2004). Exploring the use of effect sizes to evaluate the impact of different influences on child outcomes: Possibilities and limitations. In K. Elliot & I. Schagen (Eds.), *What Does it Mean? The Use of Effect Sizes in Educational Research* (pp. 6-24.). Slough: NFER.

Elliott, C. D., Smith, P., & McCulloch, K. (1996). British Ability Scales Second Edition (BAS II). Administration and Scoring Manual. London: Nelson. <https://doi.org/0123.x>

Field, A. P. (2009). *Discovering statistics using SPSS : (and sex and drugs and rock “n” roll)*. SAGE Publications.

Garner, C. L., & Raudenbush, S. W. (1991). Neighborhood Effects on Educational Attainment: A Multilevel Analysis. *Sociology of Education*, 64(4), 251. <https://doi.org/10.2307/2112706>

Green, J. A. K., & Goswami, U. (2008). Synesthesia and number cognition in children. *Cognition*, 106(1), 463–473. <https://doi.org/10.1016/j.cognition.2007.01.013>

- Gross, V. C., Nearing, S., Caldwell-Harris, C. L., & Cronin-Golomb, A. (2011). Superior Encoding Enhances Recall in Color-Graphemic Synesthesia. *Perception, 40*(2), 196–208. <https://doi.org/10.1068/p6647>
- Jäncke, L., Beeli, G., Eulig, C., & Hänggi, J. (2009). The neuroanatomy of grapheme–color synesthesia. *European Journal of Neuroscience, 29*(6), 1287–1293. <https://doi.org/10.1111/j.1460-9568.2009.06673.x>
- Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron, 67*(5), 713–727. <https://doi.org/10.1016/j.neuron.2010.08.038>
- Lee, C. S., Huggins, A. ., & Theriault, D. . (2014). A measure of creativity or intelligence? Examining internal and external structure validity evidence of the Remote Associates Test. *Psychology of Aesthetics, Creativity, and the Arts, 8*(4), 446–460.
- Mednick, A. (1968). The Remote Associates Test\*. *The Journal of Creative Behavior, 2*(3), 213–214. <https://doi.org/10.1002/j.2162-6057.1968.tb00104.x>
- Meier, B., & Rothen, N. (2013a). Grapheme-color synaesthesia is associated with a distinct cognitive style. *Frontiers in Psychology, 4*, 632. <https://doi.org/10.3389/fpsyg.2013.00632>
- Meier, B., & Rothen, N. (2013b). Synesthesia and Memory. In Julia Simner & E. M. Hubbard (Eds.), *The Oxford Handbook of Synesthesia* (pp. 692–707). Oxford: Oxford University Press.
- Meier, B., Rothen, N., & Walter, S. (2014). Developmental aspects of synaesthesia across the adult lifespan. *Frontiers in Human Neuroscience, 8*, 129. <https://doi.org/10.3389/fnhum.2014.00129>
- Melhuish, E. (2010). Impact of the Home Learning Environment on child cognitive

development: Secondary analysis of data from “Growing Up in Scotland.” *Faculty of Social Sciences - Papers*.

Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Developmental Science*, 6(2), 136–142. <https://doi.org/10.1111/1467-7687.00263>

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76(3), 241–263. <https://doi.org/10.1037/h0027272>

Pritchard, J., Rothen, N., Coolbear, D., & Ward, J. (2013). Enhanced associative memory for colour (but not shape or location) in synaesthesia. *Cognition*, 127(2), 230–234. <https://doi.org/10.1016/J.COGNITION.2012.12.012>

Radvansky, G. A., Gibson, B. S., & McNeerney, M. W. (2011). Synesthesia and memory: Color congruency, von Restorff, and false memory effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 219–229. <https://doi.org/10.1037/a0021329>

Ramachandran, V. S., & Hubbard, E. M. (2001). Psychophysical investigations into the neural basis of synaesthesia. *Proceedings of the Royal Society B: Biological Sciences*, 268(1470), 979–983. <https://doi.org/10.1098/rspb.2000.1576>

Ramachandran, Vilayanur S., & Azoulay, S. (2006). Synesthetically Induced Colors Evoke Apparent-Motion Perception. *Perception*, 35(11), 1557–1560. <https://doi.org/10.1068/p5565>

Raviselvam, S., Anderson, D., Holtta-Otto, K., & Wood, K. L. (2018). Systematic framework to apply extraordinary user perspective to capture latent needs among ordinary users. In *Proceedings of the ASME international design engineering technical conferences and*

*computers and information in engineering conference, 2018, VOL 7.*

- Rich, A. N., Bradshaw, J. L., & Mattingley, J. B. (2005). A systematic, large-scale study of synaesthesia: implications for the role of early experience in lexical-colour associations. *Cognition*, 98, 53–84. <https://doi.org/10.1016/j.cognition.2004.11.003>
- Rothen, N., & Meier, B. (2010). Grapheme–colour synaesthesia yields an ordinary rather than extraordinary memory advantage: Evidence from a group study. *Memory*, 18(3), 258–264. <https://doi.org/10.1080/09658210903527308>
- Rothen, N., Meier, B., & Ward, J. (2012). Enhanced memory ability: Insights from synaesthesia. *Neuroscience & Biobehavioral Reviews*, 36(8), 1952–1963. <https://doi.org/http://dx.doi.org/10.1016/j.neubiorev.2012.05.004>
- Rothen, N., Seth, A. K., Witzel, C., & Ward, J. (2013). Diagnosing synaesthesia with online colour pickers: Maximising sensitivity and specificity. *Journal of Neuroscience Methods*, 215(1), 156–160. <https://doi.org/http://dx.doi.org/10.1016/j.jneumeth.2013.02.009>
- Rouw, R., & Scholte, H. S. (2007). Increased structural connectivity in grapheme-color synesthesia. *Nature Neuroscience*, 10(6), 792–797. <https://doi.org/10.1038/nn1906>
- Sammons, P., Great Britain. Department for Education., J., Smees, R., Goff, J., Sylva, K., Smith, T., ... Smith, G. (2015). *The impact of children's centres : studying the effects of children's centres in promoting better outcomes for young children and their families : evaluation of children's centres in England (ECCE, strand 4)*. Department for Education.
- Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., & Hunt, S. (2008). *Effective Pre-school and Primary Education 3-11 Project (EPPE 3-11) Influences on*

*Children's Attainment and Progress in Key Stage 2: Cognitive Outcomes in Year 6.*

London: DfE.

ScotCen Social Research. (2018). Data documentation. Retrieved September 11, 2018, from

<https://growingupinScotland.org.uk/using-gus-data/data-documentation/>

Scotland. Scottish Executive. (2007). *Growing up in Scotland : a study following the lives of Scotland's children*. Scottish Executive.

Scottish Credit and Qualifications Framework. (2006).

Simner, J., & Bain, A. E. (2013). A longitudinal study of grapheme-color synesthesia in childhood: 6/7 years to 10/11 years. *Frontiers in Human Neuroscience*, (NOV).

<https://doi.org/10.3389/fnhum.2013.00603>

Simner, J., Glover, L., & Mowat, A. (2006). Linguistic determinants of word colouring in grapheme-colour synaesthesia. *Cortex*, 42(2), 281–289. [https://doi.org/10.1016/S0010-9452\(08\)70353-8](https://doi.org/10.1016/S0010-9452(08)70353-8)

Simner, J., Ipser, A., Smees, R., & Alvarez, J. (2017). Does synaesthesia age? Changes in the quality and consistency of synaesthetic associations. *Neuropsychologia*, 106, 407–416.

<https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2017.09.013>

Simner, J., & Bain, A. E. (2017). Do children with grapheme-colour synaesthesia show cognitive benefits? *British Journal of Psychology*. <https://doi.org/10.1111/bjop.12248>

Simner, J., Harrold, J., Creed, H., Monro, L., & Foulkes, L. (2009). Early detection of markers for synaesthesia in childhood populations. *Brain*, 132(1), 57–64.

<https://doi.org/10.1093/brain/awn292>

Simner, J., Hughes, J. A. E., Carmichael, D. A., & Smees, R. (n.d.). The MULTISENSE grapheme-colour diagnostic for adolescents.



- Simner, J, Ipser, A., Smees, R., & Alvarez, J. (2017). Does Synaesthesia Age? Changes in the quality and consistency of synaesthetic associations. *Neuropsychologica*, *106*, 407–416.
- Simner, JULIA. (2019). *SYNAESTHESIA : a very short introduction*. OXFORD UNIV Press.
- Simner, Julia. (2012). Defining synaesthesia. *British Journal of Psychology*, *103*(1), 1–15.  
<https://doi.org/10.1348/000712610X528305>
- Simner, Julia, & Logie, R. (2007). Synaesthetic consistency spans decades in a lexical-gustatory synaesthete. *Neurocase*, *13*(5–6), 358–365.  
<https://doi.org/10.1080/13554790701851502>
- Simner, Julia, Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., & Oakley, D. A. (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cognitive Neuropsychology*, *22*(8), 1069–1085.  
<https://doi.org/10.1080/02643290500200122>
- Terhune, D. B., Wudarczyk, O. A., Kochuparampil, P., & Cohen Kadosh, R. (2013). Enhanced dimension-specific visual working memory in grapheme–color synesthesia. *Cognition*, *129*(1), 123–137. <https://doi.org/10.1016/J.COGNITION.2013.06.009>
- Ward, J., Hovard, P., Jones, A., & Rothen, N. (2013). Enhanced recognition memory in grapheme-color synaesthesia for different categories of visual stimuli. *Frontiers in Psychology*, *4*, 762. <https://doi.org/10.3389/fpsyg.2013.00762>
- Ward, J., & Simner, J. (2003). Lexical-gustatory synaesthesia: Linguistic and conceptual factors. *Cognition*, *89*(3), 237–261. [https://doi.org/10.1016/S0010-0277\(03\)00122-7](https://doi.org/10.1016/S0010-0277(03)00122-7)
- Ward, J., Thompson-Lake, D., Ely, R., & Kaminski, F. (2008). Synaesthesia, creativity and art: what is the link? *British Journal of Psychology (London, England : 1953)*, *99*(Pt 1), 127–141. <https://doi.org/10.1348/000712607X204164>

- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale III*. The Psychological Corporation, San Antonio, TX. <https://doi.org/10.1177/1073191102009001003>
- Wechsler, D. (2004). *The Wechsler intelligence scale for children. The Wechsler intelligence scale for children—fourth edition*.
- Wechsler, D. (2005). Wechsler Individual Achievement Test - Second UK Edition (WIATIIUK). Oxford: Harcourt Assessment. <https://doi.org/10.1136/adc.2008.152793>
- Wilson Vanvoorhis, C. R., & Morgan, B. L. (2007). *Understanding Power and Rules of Thumb for Determining Sample Sizes. Tutorials in Quantitative Methods for Psychology* (Vol. 3).
- Yaro, C., & Ward, J. (2007). Searching for Shereshevskii: What is superior about the memory of synaesthetes? *Quarterly Journal of Experimental Psychology*, 60(5), 681–695. <https://doi.org/10.1080/17470210600785208>

*Learning in colour: Children with grapheme-colour synaesthesia show cognitive benefits in vocabulary and self-evaluated reading*

**Supplementary Information.**

	<i>Red</i>	<i>Blue</i>	<i>Green</i>	<i>Yellow</i>	<i>Orange</i>	<i>Pink</i>	<i>Purple</i>	<i>Brown</i>	<i>Black</i>	<i>Grey</i>	<i>White</i>
<i>R</i>	255	0	0	255	255	255	120	138	0	130	255
<i>G</i>	0	100	160	255	128	0	0	76	0	130	255
<i>B</i>	25	255	0	0	0	255	150	0	0	130	255

Figure 1. Figure shows the red, green, blue (RGB) coordinates for the colour patches used in our synaesthesia diagnostic consistency test. Values are taken from Berlin and Kay (1969) original RGB values for prototypical colours, with the exception of Grey and Brown which were marginally adjusted for print quality. In Berlin and Kay (1969) Grey was originally 115,115,115; and Brown was originally 110, 60, 0

**Sensitivity (power) analysis of prior results from (Simner & Bain, 2018)**

Here we consider the data collected by (Simner & Bain, 2018) to establish whether they had sufficient power to detect differences between synaesthetes and non-synaesthete controls in their receptive vocabulary task. Using the power analysis statistical program G\*Power (Faul, Erdfelder, Lang, & Buchner, 2007) we estimated sensitivity in terms of the size of effect required with the sample sizes (n=5). Simner & Bain (2018) used a Wilcoxon signed rank one sample test to assess any significant differences between a population mean (M = 100), and their sample mean. At .80 power (p<.05 significance, two tailed significance), a sample size of 5 would require an Effect size of 1.8, to get a significant effect (actual ES = 0.7).

### Models split by synaesthesia type

		<i>Receptive Vocabulary</i>	
		Estimate (SE)	d (95% CI)
Intercept		10.77 (0.22)	
Age~		0.06 (0.02)**	0.19 (0.08,0.29)
Gender: Boys (vs Girls)		0.75 (0.09)**	0.43 (0.31,0.52)
Highest household qualification (vs degree)	None	-1.76 (0.29) **	-1.01 (-1.35,-0.58)
	Other	-0.17 (0.73)	-0.09 (-0.71,0.55)
	Lower	-0.66 (0.29)*	-0.38 (-0.71,-0.01)
	Upper	-0.80 (0.14)**	-0.46 (-0.65,-0.24)
	Higher	-0.20 (0.11)#	-0.12 (-0.21,-0.02)
Number only (vs high memory)		1.00 (0.56)*	0.58 (0.06,0.99)
Letter only (vs high memory)		0.19 (0.45)	0.11 (-0.42,0.57)
Number and Letter (vs high memory)		0.78 (0.46)	0.45 (-0.19,0.99)
Average memory (vs high memory)		-0.30 (0.21)	-0.17 (-0.40,0.02)
Random: PSU		0.21 (0.08)	
Random: Child		3.04 (0.12)**	
N		1531	

**SI Table 1:** Mixed Linear effects models for Receptive Vocabulary at Sweep 8 (parentheses show standard error) comparing three synaesthesia sub-types to average memory controls. \* $p < 0.05$  \*\* $p < 0.01$  \*\*\*  $p < 0.001$ , ~grand centred mean, nf not fitted. d Cohen's d (size of effect only); Bootstrapped significance and CI values shown

		<i>Sentence Comprehension</i>	
		B (SE)*	OR (95% CI)
Intercept		-0.41 (0.27)***	
Age~		nf	
Gender: Boys (vs Girls)		-0.39 (0.12)**	0.68 (0.53,0.86)
Highest household qualification (vs degree)	None	-0.71 (0.46)	0.49 (0.20,1.21)
	Other	0.06 (0.90)	1.06 (0.18,6.20)
	Lower	-0.67 (0.46)	0.51 (0.21,1.27)
	Upper	-0.18 (0.18)	0.84 (0.59,1.19)
	Higher	-0.34 (0.14)*	0.71 (0.54,0.94)
Number only (vs high memory)		0.66 (0.65)	1.94 (0.54,6.97)
Letter only (vs high memory)		-0.74 (0.63)	0.48 (0.14,1.65)
Number and Letter (vs high memory)		0.16 (0.56)	1.17 (0.39,3.51)
Average memory (vs high memory)		-0.30 (0.27)	0.74 (0.44,1.25)
Random: PSU		0.26 (0.09)**	
Random: Child		---	
N		1531	

**SI Table 2:** Mixed Binary logistic effects for Sentence Comprehension at Sweep 8 (parentheses show standard error) comparing three synaesthesia sub-types to average memory controls. \* $p < 0.05$  \*\* $p < 0.01$  \*\*\*  $p < 0.001$ , ~grand centred mean, nf not fitted. d Cohen's d

		<i>Expressive Vocabulary</i>	
		<i>Estimate (SE)</i>	<i>d (95% CI)</i>
Intercept		8.60 (0.26)***	
Age~		0.10 (0.02)***	0.25 (0.15,0.35)
Gender: Boys (vs Girls)		nf	
Highest household qualification (vs degree)	None	-1.72 (0.35)***	-0.84 (-1.17,-0.51)
	Other	-1.27 (0.86)	-0.62 (-1.45,0.20)
	Lower	-0.80 (0.35)*	-0.39 (-0.72,-0.06)
	Upper	-0.78 (0.16)***	-0.38 (-0.54,-0.23)
	Higher	-0.50 (0.12)***	-0.24 (-0.36,-0.13)
Number only (vs high memory)		1.80 (0.65)**	0.88 (0.25,1.50)
Letter only (vs high memory)		0.42 (0.53)	0.21 (-0.30,0.71)
Number and Letter (vs high memory)		1.01 (0.54)#	0.49 (-0.02,1.01)
Average memory (vs high memory)		-0.32 (0.25)	-0.16 (-0.40,0.08)
Random: PSU		0.38 (0.10)***	
Random: Child		4.19 (0.16)***	
N		1531	

**SI Table 3:** Mixed Linear effects models/Binary logistic effects for Expressive Vocabulary at Sweep 8 (parentheses show standard error) comparing three synaesthesia sub-types to average memory controls. \*p<0.05 \*\*p<0.01 \*\*\* p<0.001, ~grand centred mean, nf not fitted. d Cohen's d

		<i>Academic Self-Concept: Reading</i>		<i>Academic Self-Concept: Number work</i>	
		Estimate (SE)	OR (95% CI)	Estimate (SE)	OR (95% CI)
Intercept		-0.30 (0.25)		-0.62 (0.25)*	
Age~		Nf		nf	
Gender: Boys (vs Girls)		Nf		0.75 (0.10)***	2.12 (1.73,2.60)
Highest household qualification (vs degree)	None	-0.96 (0.39)*	0.38 (0.18,0.82)	nf	
	Other	0.09 (0.83)	1.09 (0.21,5.61)	nf	
	Lower	-0.85 (0.38)*	0.43 (0.20,0.89)	nf	
	Upper	-0.75 (0.17)***	0.47 (0.34,0.66)	nf	
	Higher	-0.31 (0.12)**	0.73 (0.58,0.93)	nf	
Number only (vs high memory)		0.33 (0.63)	1.38 (0.40,4.74)	-0.17 (0.64)	0.84 (0.24,2.97)
Letter only (vs high memory)		1.04 (0.53)*	2.83 (1.00,7.99)	-0.14 (0.52)	0.87 (0.31,2.42)
Number and Letter (vs high memory)		1.00 (0.54)#	2.72 (0.94,7.84)	-0.07 (0.53)	0.93 (0.33,2.64)
Average memory (vs high memory)		0.06 (0.25)	1.06 (0.65,1.73)	-0.02 (0.25)	0.98 (0.61,1.59)
Random: PSU n		nf 1531		nf 1531	

**SI Table 4:** Binary logistic effects for Academic self-concept in Reading and Number Work at Sweep 8 (parentheses show standard error) comparing three synaesthesia sub-types to average memory controls.\* $p < 0.05$  \*\* $p < 0.01$  \*\*\*  $p < 0.001$ , ~grand centred mean, nf not fitted. d Cohen's d AM Average memory

		<i>Expressive Vocabulary</i>	
		Coefficient (SE)	d (95% CI)
Intercept		64.04 (1.21)***	
Age~		nf	
Gender: Boys (vs Girls)		-1.42 (0.49)**	-0.15 (-0.26,-0.05)
Highest household qualification (vs degree)	None	-10.36 (1.53)***	-1.11 (-1.43,-0.79)
	Other	-2.30 (3.90)	-0.25 (-1.07,0.57)
	Lower	-6.53 (1.56)***	-0.70 (-1.03,-0.37)
	Upper	-4.20 (0.73)***	-0.45 (-0.60,-0.30)
	Higher	-2.67 (0.57)	-0.29 (-0.41,-0.17)
Synaesthete (vs high memory)		0.53 (1.76)	0.06 (-0.31,0.43)
average memory (vs high memory)		-1.50 (1.17)	-0.16 (-0.41,0.09)
Random: PSU		3.77 (1.52)*	
Random: Child		87.25 (3.35)***	
n		1486	

**SI Table 5.** Mixed effects Linear regression model for Expressive vocabulary at Sweep 5 (parentheses show standard error). \* $p < 0.05$  \*\* $p < 0.01$  \*\*\*  $p < 0.001$ , ~grand centred mean. Synaesthetes and average memory controls are compared to high memory controls. d Cohen's d (size of effect only)

## References

- Berlin, B., & Kay, P. (1969). *Basic color terms : their universality and evolution*. University of California Press.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. <http://doi.org/10.3758/BF03193146>
- Simner, J., & Bain, A. E. (2018). Do children with grapheme-colour synaesthesia show cognitive benefits? *British Journal of Psychology*, *109*(1). <http://doi.org/10.1111/bjop.12248>