

Performance on the Farnsworth-Munsell 100-Hue Test Is Significantly Related to Nonverbal IQ

Matthew B. Cranwell, Bradley Pearce, Camilla Loveridge, and Anya C. Hurlbert

Institute of Neuroscience, Newcastle University, Newcastle Upon Tyne, United Kingdom

Correspondence: Matthew B. Cranwell, Institute of Neuroscience, Henry Wellcome Building, Newcastle University, Framlington Place, Newcastle Upon Tyne, UK, NE2 4HH; m.b.cranwell@ncl.ac.uk

Submitted: November 18, 2014

Accepted: March 16, 2015

Citation: Cranwell MB, Pearce B, Loveridge C, Hurlbert AC. Performance on the Farnsworth-Munsell 100-Hue test is significantly related to nonverbal IQ. *Invest Ophthalmol Vis Sci.* 2015;56:3171–3178. DOI:10.1167/iovs.14-16094

PURPOSE. The Farnsworth-Munsell 100-Hue test (FM100) is a standardized measure of chromatic discrimination, based on colored cap-sorting, which has been widely used in both adults and children. Its dependence on seriation ability raises questions as to its universal suitability and accuracy in assessing purely sensory discrimination. This study investigates how general intellectual ability relates to performance on both the FM100 and a new computer-based chromatic discrimination threshold test, across different age groups in both typical and atypical development.

METHODS. Participants were divided into two main age groups, children (6–15 years) and young adults (16–25 years), with each group further subdivided into typically developing (TD; three groups; TD 6–7 years, TD 8–9 years, TD Adult) individuals and atypically developing individuals, all but one carrying a diagnosis of Autism Spectrum Disorders (ASD; two groups; atypically developing [ATY] child 7–15 years, ASD Adult). General intelligence was measured using the Wechsler Abbreviated Intelligence Scale and Wechsler Intelligence Scale for Children. All participants completed the FM100. Both child groups also completed a computer-based chromatic discrimination threshold test, which assessed discrimination along cone-opponent (“red-green,” “blue-yellow”) and luminance cardinal axes using a controlled staircase procedure.

RESULTS. Farnsworth-Munsell 100-Hue test performance was better in adults than in children. Furthermore, performance significantly positively correlated with nonverbal intelligence quotient (NVIQ) for all child groups and the young adult ASD group. The slope of this relationship was steeper for the ASD than TD groups. Performance on the chromatic discrimination threshold test was not significantly related to any IQ measure. Regression models reveal that chromatic discrimination threshold, although a significant predictor of FM100 performance when used alone, is a weaker predictor than NVIQ used alone or in combination.

CONCLUSIONS. The results indicate that FM100 performance is not purely a measure of color discrimination but instead also reflects general nonverbal ability. Other measures of chromatic discrimination ability are therefore required for its accurate assessment, particularly in early or atypical development.

Keywords: color vision assessment, visual development, color vision

Color perception is a major contributing factor to a wide range of different behavioral tasks, including, for example, visual search,¹ object recognition,² and evaluation of material properties (e.g., ripeness of fruit or healthiness of skin³). The low-level ability to discriminate between colors (chromatic discrimination ability) also underlies development of higher-level abilities and behaviors such as color naming,⁴ affective responses to color,⁵ and color memory.⁶ In recent years, color perception has been increasingly studied in developmental disorders such as autism spectrum disorder (ASD),^{7–9} Attention Deficit Hyperactivity Disorder,^{10,11} and Williams Syndrome,¹² as the extent of atypical sensory processing across all visual domains has become more evident in these disorders. It is therefore increasingly important to ensure that the diagnostic tools used to assess sensory processing, and in particular color perception, are both sensitive and specific in isolating atypicalities in sensory processing. This requirement is complicated by the hypothesized relationship between sensory

processing and intelligence^{13,14}; according to the original hypothesis of Galton and Spearman,^{13–15} higher intelligence is associated with better sensory discrimination abilities. The support for this hypothesis has been mixed,^{15–19} with low correlations found between general intelligence and some measures of sensory discrimination (including color perception), and other more recent evidence¹⁹ demonstrating a strong link between IQ and performance on a visual motion discrimination task. Nonetheless, the putative relationship makes it vital to ensure that tests of sensory processing are not confounded by direct contributions of general ability to performance.

The Farnsworth-Munsell 100-Hue Test²⁰ has been widely used by clinicians and visual scientists as a measure of chromatic discrimination ability²¹ in both typically developing individuals with normal color vision,²² and in individuals with congenital or acquired color vision deficits^{23–26} or developmental disorders.^{7,8,10,12} The FM100 test involves arranging a

set of individual colored caps of similar lightness and saturation in order between the hues of two fixed caps (e.g., blue and green), so that a smooth color gradient is formed, with the hue differences between neighboring caps as small as possible. The FM100 has been used in a wide range of ages, from early childhood (5 years) to elderly, and has the advantage of age-expected norms.^{22,27} It is also useful in identifying congenital and acquired retinal diseases and as a measure of lens yellowing during normal aging.²⁸

Other laboratory measurements of chromatic discrimination ability,^{29,30} for example, using detection tasks for forms defined by chromatic contrast embedded in backgrounds with dynamic random luminance variation, conclude that chromatic discrimination ability changes over the life span with sensitivity peaking in late adolescence/the twenties^{31,32} and subsequently decreasing throughout the remaining adulthood.^{22,27,33} The improvement in FM100 performance with age up to early adulthood, evident in studies that establish age-dependent norms,^{22,27} is consistent with this age-related enhancement of chromatic discrimination ability. Yet, other studies suggest that this change in FM100 performance may be due at least partly to improvements in the ability to perform seriation tasks.^{34,35} A recent study concludes that the general ability to perform a seriation task has sufficiently developed by the age of 5 that its further development does not explain the improvement in FM100 with age, but also that there are other noncolor-specific factors related to the difficulty of the discrimination which affect FM100 performance.³⁶ It has also been observed that the Roth-28 (a shortened version of the FM100 using every third cap) is a time-consuming test that requires attention and a degree of visuomotor competence, and is subject to learning and practice effects,³⁷ and therefore the question has been raised as to whether other tests of color discrimination might be more useful or appropriate.³⁸ Although a previous study concludes that the specific correlation between the FM100 and general intelligence is low,¹⁶ we have found (in a preliminary study to that reported here) that performance on the FM100 is related to nonverbal ability (NVIQ) in adults,⁹ and more strongly so in individuals diagnosed with ASD. Taken together, these results argue strongly for dissociating the various factors contributing to performance on the FM100, including age, attentional ability, general intelligence, and chromatic discrimination ability, before accepting it as an unequivocal measure of chromatic discrimination ability suitable for widespread diagnostic use.

Here, the contribution of general cognitive ability to performance on two chromatic discrimination tests (the FM100 and a computer-based form-discrimination threshold task) is assessed. Specifically, we aim to determine the extent to which the relationship between nonverbal general ability and performance on the FM100 varies with age (e.g., between adults and children) and also with development, between typical and atypical development.

METHODS

Participants

Ninety-two participants took part in the study, split across five different participant groups on the basis of chronological age (adult/child) and the typicality of their development (typically developing [TD]/atypically developing [ATD]). All adult participants gave informed consent. For the child groups, parental consent was obtained and assent was given by the participant prior to the start of the experiment in line with the Declaration of Helsinki. Ethical approval was given from the Faculty of Medical Sciences Ethics Committee, Newcastle

TABLE 1. Participant Group Demographics

Group	Chronological Age	VIQ	NVIQ
TD adult, <i>n</i> = 28	20.07 (2.04)	96.36 (10.47)	108.89 (9.23)
ASD adult, <i>n</i> = 17	18.12 (1.22)	86.35 (19.05)	93.59 (18.94)
TD 6-7 y, <i>n</i> = 11	6.37 (0.17)	111.45 (10.78)	109.27 (10.11)
TD 8-9 y, <i>n</i> = 17	8.95 (0.31)	121.24 (11.84)	112.24 (13.79)
ATY child, <i>n</i> = 15	12.06 (2.48)	71.4 (22.82)	85.53 (20.21)

Table includes only those participants whose results were included in the FM100 Analysis. Chronological age is shown in years. Verbal and nonverbal IQ are shown as standardized scores. In both adult groups, IQ was assessed using the WASI. In child groups (TD 6-7 years, TD 8-9 years, and ATY child) IQ was assessed using the WISC Fourth Edition. Standard deviations are shown in parentheses (see text for explanation).

University (Approval Code 00618/2012; for child participants; Newcastle Upon Tyne, UK) and from the Newcastle University Psychology Ethics Committee (Approval Code 060004; for adult participants).

There were two adult groups: a TD group (TD adult; *n* = 28, mean age = 20.07 years, male = 15, female = 13) and an ATD group diagnosed with ASD (ASD adult; *n* = 18, mean age = 18.11, male = 15, female = 3). There were no psychiatric or developmental disorder diagnoses in the TD Adult group. Intelligence was assessed using the Wechsler Abbreviated Scale of Intelligence (WASI).³⁹ Verbal IQ (VIQ) was determined by Vocabulary and Similarities subtests, and Performance or nonverbal IQ (NVIQ) was measured using Block Design and Matrix Reasoning subtests.

Three groups of children were recruited from local mainstream and special needs schools: two TD groups, 12 children aged 6- to 7-years old (male = 5, female = 7) and 19 children aged 8- to 9-years old (male = 5, female = 14); and one group of 17 children diagnosed with ASD plus one child with Williams Syndrome (ATY child; male = 12, female = 6). There were no developmental disorders reported for any children in the two TD groups. Four subtests of the Wechsler Intelligence Scale for Children (WISC⁴⁰) Fourth Edition were administered as measures of IQ: block design, picture concepts, vocabulary, and similarities. Nonverbal IQ was determined by the combined scores on block design and picture concepts subtests. Verbal IQ was calculated from the scores on vocabulary and similarities subtests (see Table 1 for included participant demographics). All participants were screened for color deficiencies using the Neitz Test of Color Vision⁴¹; all included participants were classified as normal trichromats.

All participants were tested on the IQ tests described above and the Farnsworth-Munsell 100-Hue test (FM100) described below. Additionally, the three child groups (TD 6-7 years, TD 8-9 years, and ATY child) were tested on the chromatic contrast discrimination threshold test, described below. This test was developed for a children's color perception test battery partly in response to the observed relationship between performance on the FM100 and nonverbal ability (NVIQ) in adults,⁹ and was therefore not available at the time of the adult testing.

Results from six participants were removed from the FM100 analysis, for the following reasons: one ASD adult, one TD 6- to 7-year-old child and two ATY child participants due to total error scores (TES) of over 500, implying poor task comprehension²²; one TD 8- to 9-year-old child and one ATY child due to task noncompletion (time constraints); one TD 8- to 9-year-old child for deuteranomaly, as evidenced by Neitz Test screening and the FM100 error pattern. Table 1 reports indices for the remaining participants, whose FM100 results were

included in analysis. On the chromatic contrast discrimination threshold test (see Supplementary Table S1), results were excluded from two ATY child participants due to task noncompletion (time constraints), and from the one TD 8- to 9-year-old child due to deuteranomaly.

Farnsworth-Munsell 100-Hue Test (FM100)

The Farnsworth-Munsell 100-Hue test²⁰ is a measure of chromatic discrimination. It consists of 85 colored caps split across four trays. The caps vary only in hue, with lightness and saturation kept constant. Each tray has 21 removable intermediate caps (with the exception of the first tray where there are 22 caps) whose hues vary smoothly between those of the two fixed caps at either end. Standard administration procedures were followed: For each tray, the intermediate caps are removed from the tray and placed in a random arrangement while the participant looks away. The participant is then asked to view and place the intermediate caps in the correct order in the tray between the two fixed caps, with as little difference in hue between neighboring caps as possible. Standard prompts of, "Which color is most like the one at the end?" were used to ensure that the task was understood correctly. The trays were completed in different orders between participants. The order in which the participant placed the caps was recorded by the experimenter. The task was completed under simulated daylight illumination of color temperature 6500K (D65) produced by a VeriVide D65 "Artificial Daylight" lamp (Leicester, UK).

Standard scoring procedures were followed. Error scores for each tray position are calculated on the basis of the differences between its chosen cap and the two neighboring caps, generating a baseline score of 2 for each cap when in perfect order. Error scores for caps at the end of each tray were calculated using the neighboring cap in the same tray and the first cap of the next tray, so that all caps are considered on a continuum around the color circle. The TES is computed by first subtracting the baseline score from each tray position error score and then summing all 85 individual error scores. Specific anomalies of color vision are revealed by specific error patterns (clustering of cap transposition errors along the protan, deutan, or tritan axes).

Chromatic Contrast Discrimination Threshold Test (CCDT)

Overview. The CCDT was designed to isolate and assess discrimination within each of the cardinal chromatic mechanisms independently, similarly to the class of contemporary chromatic discrimination tests which include the Colour Assessment and Diagnosis (CAD) test²⁹ and the Cambridge Colour Test (CCT).³⁰ It was developed for use in this and related studies of color perception in children, with the requirements that it be engaging to children, portable, relatively quick to run, and easily reproduced without specialist equipment. The thresholds measured by the CCDT are comparable with those from the CAD and CCT, although exact comparisons cannot be made because of differences in the specific shape discrimination task used and the background chromaticity and luminance. The CCDT differs from the Farnsworth-Munsell 100-Hue test in that it directly measures thresholds for chromatic discrimination around a point of neutral chromaticity, whereas the FM100 does not measure thresholds but instead requires the observers to detect (and then seriate) chromatic differences between colors at a fixed distance from neutral chromaticity. These chromatic differences have been selected to be near threshold for normal observers (note that, in general, discrimination thresholds for hues of roughly equal luminance and saturation will differ from thresholds relative to a neutral chromaticity). Despite the

differences between the two types of test, other studies suggest that the age dependence for both is similar,³³ supporting the assumption that both rely on the same basic chromatic processing mechanisms.

Apparatus and Setup. Color stimuli were displayed on a computer screen placed at the back of a black viewing box (36 × 44.8 × 62.3 cm). Participants rested their heads on a chin rest placed centrally at the front of the box and viewed the screen through an aperture (13.5 × 9 cm) placed 21 cm along the box length, from a distance of 62 cm. One of two different computer setups was used to control the experiment, depending on the group: for the TD 6- to 7-year olds and ATY child groups, the experiment ran on a Dell Inspiron Laptop (Plano, TX, USA) with stimuli displayed on its 14-inch screen; while for the TD 8- to 9-year-old group the experiment ran on a custom built portable desktop tower, with standard components, running Windows 7 64-Bit edition (Reading, UK) with a PNY 600 10-bit graphics board (NVIDIA, Almondsbury, Bristol, UK) with the stimuli displayed on a 10-bit 23-inch Proart LCD monitor PA 238Q (London, UK) using a display port adapter. The same experimental programme (the Chromatic Contrast Discrimination Threshold test, or CCDT) was used for both setups, written in Matlab (v7.6.0, 2012b; The MathWorks, Cambridge, UK), with graphics display functions from Psychtoolbox⁴² and colorimetric conversion functions from *kccv* (a set of Matlab routines based on standard formulae⁴³ tailored for 8- or 10-bit displays appropriately); the 10-bit display used the NVIDIA QUADRO performance drivers (Almondsbury, Bristol, UK). Spectral emission properties of both screens were characterized using a PR-650 spectroradiometer (SpectraScan, Photo Research, Inc., CA, USA) and colorimetric calibration tables were checked regularly using a Minolta CS-100 chromameter (The Konica Minolta, Nieuwegein, Netherlands) and updated when necessary to ensure colorimetric accuracy of the displayed stimuli.

Stimuli. On each trial, a single colored arrow (visual angle = 1.83°), pointing either left or right, was presented. The vertical position of the arrow was randomly jittered from trial to trial 5.51° above and below the central fixation point (visual angle = 0.92°), on an achromatic gray background (CIE 1931 coordinates: $x = 0.36$, $y = 0.37$; $Y = 20.46$ cd/m² for the 8-bit display; $x = 0.314$, $y = 0.339$; $Y = 64.8$ cd/m² for the 10-bit display). The arrow color was systematically varied in increments along only one of the three cone-opponent-contrast axes⁴⁴ (L-M or 'Red-Green'; S-(L+M) or 'Blue-Yellow'; L+M or 'luminance'; see Supplementary Material for further detail). The just-noticeable difference in arrow color with respect to the background was calculated in ΔE units in a perceptually uniform color space (CIE L*u*v*space).

Design. A staircase protocol was used to vary the color difference between the arrow and background on each trial, stepping through differences on each half of the red-green, blue-yellow, and luminance cone-opponent-contrast axes separately, beginning with suprathreshold difference values and moving in progressively smaller increments to difference values that are just reliably detected by the observer. A one-up/two-down procedure was used, in which the participant must be correct twice consecutively to go down the staircase (i.e., testing smaller color differences), whereas an incorrect answer will take the participant up the staircase (i.e., testing larger color differences; details of step sizes are provided in the Supplementary Material). Each color axis was tested in a separate block of trials, with a maximum of 100 trials per color axis (50 for each color direction of the axis; e.g., 50 each for "bluer" and "yellower") and a maximum number of 30 reversals per half-axis, with the two independent staircases for each half-axis randomly interleaved in the one block. Each trial began with a 500 ms centrally positioned white fixation

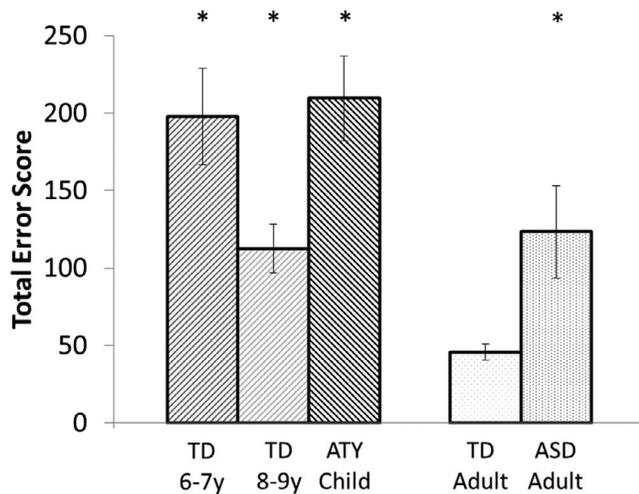


FIGURE 1. Total error scores (TES) for the FM100 for each participant group. Higher TES reflects poorer performance on the test. Errors bars are \pm one standard error. The * denotes a significant difference from the expected age-appropriate norms reported by Kinnear and Sahraie.²²

dot, followed by the target that appeared for 150 ms. The participant had to respond whether the arrow pointed to either the left or right, by pressing the corresponding mouse button. There was no time limit on the response, and the next trial began immediately after the response had been given.

Procedure. Participants sat in front of the viewing box. Prior to the start of the experiment, the researcher checked that the participant understood the difference between left and right. A short practice set of highly visible arrows (with suprathreshold luminance contrast) was administered to check that the participant understood the experiment. After the practice set the actual experiment began. In a game-like format, participants were presented with a choice for each condition (each depicted by a different storybook image) and performed each condition only once. The order of condition was not counterbalanced but there was no significant difference of order for any of the groups (lowest $P = 0.11$). Each color-axis condition ran until either the maximum number of trials or maximum number of staircase reversals was reached. Once each condition had finished the participant was given a short break before continuing to the next condition. All conditions were either completed in one session or over two sessions.

RESULTS

Experiment 1: Farnsworth-Munsell 100-Hue Test Results

The error patterns of all included participants revealed no specific anomalies of color vision. A one-way ANOVA conducted with TES as the dependent variable and a between-subjects factor of participant group found a significant difference in TES between groups ($F(4, 83) = 11.83, P < 0.001$; Fig. 1). Subsequent post hoc t -tests using a Bonferroni-corrected P value revealed specific group differences: The TD adult group had significantly lower TES (mean TES = 45.79) than all other groups (highest $P < 0.001$) and the TD 8- to 9-year-old group had significantly lower TES (mean TES = 112.47) than the ATY child group (mean TES = 209.47), ($t(30) = 3.16, P < 0.05$) (TD 6- to 7-year-old group mean TES = 197.64; ASD adult group mean TES = 123.29). No other

specific group TES comparisons revealed significant differences (lowest $P = 0.115$).

Correlation of IQ With TES

Correlations were calculated for each group between TES and standardized scores on VIQ and NVIQ subscales of respective IQ tests. For VIQ and TES correlations, significant correlations were found for ASD adult ($r = -0.68, P < 0.005$), and ATY child groups ($r = -0.58, P < 0.05$). No other VIQ/TES correlations were significant (lowest $P = 0.33$). For NVIQ correlations with TES there were significant correlations for ASD adults ($r = -0.89, P < 0.001$), TD 6- to 7-year olds ($r = -0.63, P < 0.05$), TD 8- to 9-year olds ($r = -0.65, P < 0.005$), and ATY child groups ($r = -0.62, P < 0.05$; Fig. 2). A trend toward significance was observed for the TD adult group ($r = -0.32, P = 0.095$).

Relation to Previously Reported Age-Norms

Comparisons between the TES found in this sample and expected age-norms²² were made. Difference scores between observed and expected scores for every participant were calculated, with a negative score indicating worse performance relative to the expected age score and a positive score better performance (Fig. 1). One-sample t -tests were conducted against a value of 0 to reveal any significant deviations in each participant sample relative to expected age norms. Significantly better than expected performance was found in the TD 6- to 7-year olds ($t(10) = 2.81, P < 0.05$) and TD 8- to 9-year olds ($t(16) = 2.4, P < 0.05$) groups and significantly worse than expected performance for the ASD adult ($t(16) = 2.65, P < 0.05$) and ATY child ($t(14) = 3.65, P < 0.005$) groups. No significant differences with age-expected norms were found for the TD adult group ($t(27) = 1.02, P = 0.32$).

Discrimination Thresholds on Chromatic and Luminance Axes

Individual staircases for each color half-axis were analyzed for convergence and excluded from further analysis if either of two conditions were met: (1) no reversals in the final 20% of trials, or (2) final stimulus contrast less than or equal to zero, relative to background chromaticity. If the staircase for one half-axis did not meet these conditions, the staircase in the other half-axis was also discarded from analysis. For any one axis (red-green [R-G], blue-yellow [B-Y], or luminance [LUM]), a maximum of six participants' staircases were excluded on these criteria for the ATY child group, and a maximum of three were excluded for the TD 6- to 7-year-old child group. All staircases in the TD 8- to 9-year-old group were accepted. All participants described in Supplementary Table S1 contributed acceptable data for at least one axis.

Threshold contrasts for each of the six half-axes were calculated in ΔE_{uv} units, as described above. The two thresholds for the two directions for each condition were then averaged together to give an overall threshold for each cardinal axis (e.g., blue and yellow individual thresholds were averaged together for a B-Y axis threshold). To normalize the threshold distributions for each color axis, ΔE_{uv} thresholds were converted to a logarithmic scale.

To compare thresholds between color axes, data collected from the 8-bit display (TD 6-7 years and ATY child groups) and 10-bit display (TD 8-9 years child group) systems were analyzed separately (see Fig. 3), given the difference in background luminance between the systems, which affects the absolute magnitude of the thresholds. For the former, a two-way ANOVA with participant group (TD 6-7 years/ATY child) and color axis as fixed factors and threshold (ΔE_{uv}) as

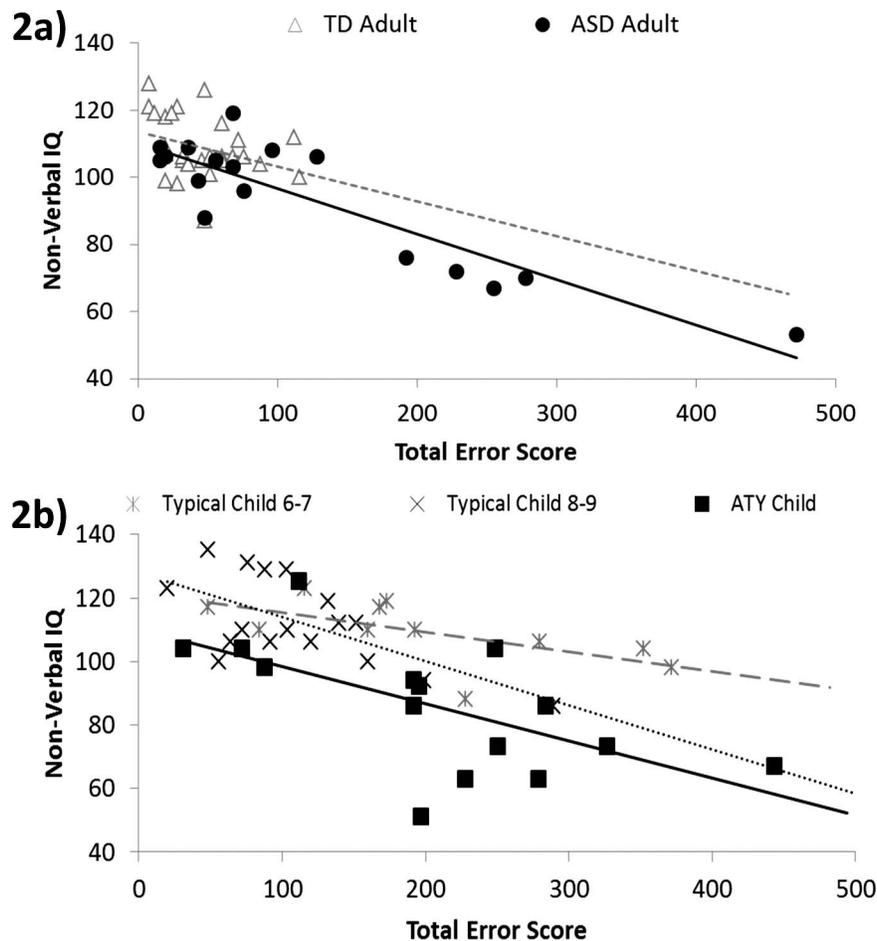


FIGURE 2. Individual standardized NVIQ scores plotted against FM100 TES. Lines are least-squares best fits for each group. In both graphs, ASD groups are represented by *filled shapes* and *solid lines*. (a) Adult groups: The TD adult group are denoted by *unfilled triangles* and the ASD adult group by *filled black circles*. (b) Child groups: The TD 6- to 7-year-old group are denoted by *gray asterisks*, the TD 8- to 9-year-old group by *gray crosses*, and the ATY child group by *filled black squares*.

dependent factor revealed a significant main effect for color axis ($F(1,56) = 15.78, P < 0.001$). Post hoc tests revealed that thresholds were significantly lower for the luminance than both R-G and B-Y color axes for both TD 6 to 7 years (highest $P < 0.001$) and ATY child groups (highest $P < 0.001$). There was no significant difference between R-G and B-Y thresholds for either group (lowest $P = 0.329$). A significant main effect was also observed for group ($F(1,56) = 15.781, P < 0.001$). Post hoc tests conducted for each color axis further revealed significant differences between groups on the B-Y axis ($t(19) = 2.93, P < 0.05$), with thresholds significantly higher for the ATY versus TD 6- to 7-year-old child groups, but not for luminance ($P = 0.22$) or R-G axes ($P = 0.19$).

Similarly, for data from the TD 8- to 9-year-old child group, a two-way repeated-measures ANOVA with a within-subjects factor of color axis (three levels) revealed a significant main effect of axis ($F(2,34) = 39.46, P < 0.001$), with luminance thresholds significantly lower than both R-G ($t(17) = 12.9, P < 0.001$), and B-Y ($t(17) = 6.18, P < 0.001$), thresholds. Red-green and B-Y thresholds did not significantly differ from each other ($t(17) = 1.6, P = 0.13$).

Correlations of Thresholds With IQ

Correlation analyses were conducted to reveal relationships between VIQ and NVIQ and discrimination thresholds on the

three color axes, using adjusted P values to control for multiple comparisons. No significant correlations were found in any participant group for either VIQ (lowest $P = 0.18$) or NVIQ (lowest $P = 0.22$).

Correlations Between FM100 and Chromatic Threshold

Correlation analyses were conducted to examine the relationship between performance on the FM100 and CCDT tests. Performance was again split via participant group. For the CCDT test, the mean chromatic discrimination threshold was calculated by averaging the B-Y and R-G thresholds. No significant correlation was found for the TD 6- to 7-year-old child group ($P = 0.73$). A significant correlation for the ATY child group was found ($r = 0.83, P < 0.001$), while a trend toward significance was revealed for the TD 8- to 9-year-old child group ($r = 0.44, P = 0.08$).

Regression Analysis

To further assess the role of possible factors on FM100 performance, a multiple regression was carried out to estimate the extent to which performance on the FM100 might be predicted by NVIQ, chronological age, and chromatic discrimination (as measured by the average B-Y and R-G thresholds

TABLE 2. Backward Stepwise Regression Model for the Contributions to FM100 Performance of Distinct Factors: NVIQ, Chromatic Discrimination, Chronological Age, and Development Typicality

Variable	Model 1			Model 2		
	<i>B</i>	SE <i>B</i>	β	<i>B</i>	SE <i>B</i>	β
NVIQ	-3.629	0.953	-0.737*	-3.743	0.859	-0.76†
Chromatic discrimination	236.143	106.66	0.369‡	231.364	103.896	0.361‡
Chronological age	-1.9	0.769	-0.557‡	-1.767	0.614	-0.518†
Development	13.69	46.23	0.064			
<i>R</i> ²		0.532			0.531	
<i>F</i> value		8.539			11.701	

Model 1 includes all of the predictor variables; Model 2 omits only development typicality.

* $P < 0.001$.

† $P < 0.01$.

‡ $P < 0.05$.

from the CCDT test. This was done to gain an overall measure of chromatic discrimination to make this analogous to the TES on the FM100). Four predictors were included in the model: NVIQ, chronological age, development typicality (typical/atypical), and chromatic discrimination threshold. The analysis included only participants who completed both the FM100 and the CCDT test, from all child groups. Predictors were entered into the regression model using the backward entry method, appropriate in the absence of an a priori theory for which predictors would explain the most variance in FM100 performance. From this procedure, two models were generated. Model 1 included all of the variables, while Model 2 included all variables except for participant group (development typicality) because this factor was found not to be a significant predictor (Table 2). Overall, both models significantly explained substantial variance in FM100 performance: Model 1 explained 53.2% of the variance ($F(4,30) = 8.54$, $P < 0.001$) while Model 2 explained 53.1% of the variance ($F(3,31) = 11.70$, $P < 0.001$). Additional regression models run on data from all participant groups alone and in combination are described in the Supplementary Material; in all models, NVIQ is always the highest predictor, and although CCDT thresholds are significant predictors when used solely, they explain much less variance than NVIQ-only models.

DISCUSSION

The major finding in this study is that nonverbal general ability differentially affects participants' performance on two different chromatic discrimination tasks. The results from experiment 1 show that performance on the FM100 is significantly associated with nonverbal ability in all groups except typically developing adults. Furthermore, this association is stronger in children, and even stronger in atypically developing individuals (the majority with ASD). The results from experiment 2 demonstrate, conversely, that this association between general ability and color perception in the younger age groups does not hold for the CCDT test.

The results therefore strongly suggest that the FM100 is not an unequivocal measure of color discrimination but conflates this with a measure of "nonverbal ability," providing evidence for previous suggestions that cognitive factors unrelated to chromatic discrimination ability influence performance on the task.^{9,36} The FM100 performance-NVIQ correlation slopes are steeper and more significant in the child groups than adult. Furthermore, the results of the regression model suggest that nonverbal ability is a significant predictor of performance in addition to age and the ability to discriminate between colors.

When considering why the FM100 is associated with general cognitive ability but the chromatic discrimination threshold test is not, the relative task demands are important. Successful performance on the FM100 requires attentional and visuospatial abilities in addition to chromatic discrimination ability. Spatial comparisons are required between the selected and nonselected caps. Attention switching between the local field, in making a comparison between two adjacent caps, and the global field, in overseeing the entire color gradient, is also essential for good performance. These task demands may be influenced by different factors in each group. In typically developing children competency of global processing does not develop until late into childhood.⁴⁵ Individuals with ASD are more likely to process visual information locally rather than globally⁴⁶ and to have difficulty switching between local and global processing.⁴⁷ Nonverbal ability may also differentially affect performance between groups. Better performance in the older TD groups may reflect more mature global processing competency than in the younger TD groups. Poorer performance in the ASD groups may be the result of difficulty both in sustaining and switching of attention between local and global fields of the FM100.

These results have implications for the FM100 norms that have previously been reported.²² These norms implicitly assume that performance is unrelated to IQ and that task demands are consistent between different ages. In the current study, both TD child groups are above average in nonverbal ability and perform significantly better than expected from the Kinnear and Sahraie²² norms. Given that the number of participants in both studies is similar for each respective age group, this comparison calls into question the reliability of the norms for these age groups and more so for the younger age groups for which the number of observers is even smaller (e.g., 9 observers for age 5 years).²² These results, combined with those from a preliminary study by Hurlbert and colleagues⁹ instead suggest that caution must be exercised when using FM100 norms, for all ages, but especially for younger children or clinical populations where the NVIQ is lower than average with respect to chronological age, given that the relationship between FM100 performance and NVIQ is stronger for groups of the latter type.

In comparison to the FM100, the CCDT has fewer task demands. This test of chromatic discrimination requires attention on a trial-by-trial basis only to identify the direction of an arrow. Like other standardized tests in whose class the CCDT falls (e.g., the CCT³⁰ and the CAD²⁹), the CCDT measures discrimination thresholds along isolated chromatic directions away from a fixed adaptation point. Although other color discrimination tasks have been adapted for use with

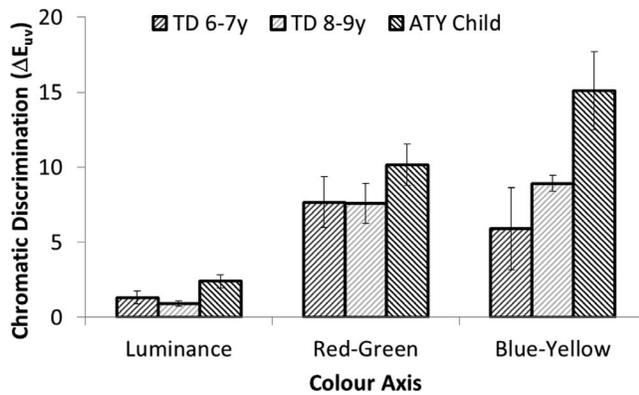


FIGURE 3. Average discrimination thresholds for the luminance, red-green and blue-yellow color axes, as determined by the CCDT test. The three groups (TD 6-7 years, TD 8-9 years, and ATY child) are indicated by different *diagonal line shading*, as shown in the *legend*. Error bars indicate standard errors.

children (e.g., the CCT⁴⁸), to our knowledge these have yet to be demonstrated as independent of general ability. Other standardized chromatic discrimination threshold tests call on more complex aspects of visual processing which may introduce additional confounds when used in children: The CCT,³⁰ for example, requires participants to identify a global shape composed from local elements while the CAD (or City Colour Vision Test⁵⁰) requires participants to discriminate the direction of a moving stimulus.^{29,49} The former thus presents a potential confound in distinguishing between deficits in local chromatic discrimination versus global shape processing, while the latter may be unable to dissociate between deficits in motion direction discrimination versus chromatic discrimination, which are known to develop at different rates in children.^{33,50} The CCDT task used in this study may provide a more direct measurement of chromatic discrimination by being a simple shape identification task, which requires only a coarse binary judgement of left versus right, does not depend on numeracy or literacy skills, and does not involve a trade-off between local and global processing, involving the discrimination of only a single large shape against a uniform background. Performance is more likely to be independent of developmental stage or cognitive ability, allowing for age variations in chromatic discrimination to be more accurately captured. Because participants continue the task only until they reach their own individual threshold, task difficulty also remains constant between individuals even though other factors such as chronological age or chromatic discrimination ability may differ between participants. Because of their shared properties, we would expect CCDT performance to share the same pattern of age dependence as that demonstrated for the CCT and CAD,³² and our ongoing studies support this expectation.

Although we have here carried out the CCDT only in the younger age groups and therefore demonstrated this independence only for those groups, we would expect the independence to hold for the adult groups also, particularly given their better chromatic discrimination ability, higher absolute IQ, and increased attentional capacity relative to children.⁵¹

It is not possible from this study to address specifically whether sensory processing is related to general ability, as suggested by other studies.^{13,19} Nonetheless, the results here do highlight the importance of using multiple tests within the same sensory processing domain when assessing this link. For example, the results of the FM100 alone would suggest that there is a link between sensory (chromatic) discrimination and

intelligence, while the results of the CCDT test alone would suggest that there is no such relationship. The differential correlation between nonverbal ability and the FM100/CCDT test demonstrate that different tests measuring the same sensory (chromatic) discrimination may give different results depending on the different task demands. In order to more accurately tease apart the possible relationship between intelligence and sensory discrimination, comparisons need to be made on performance from multiple tests measuring the same sensory domain.

In summary, we show that general ability affects performance on two different color discrimination tasks differently: performance on the FM100 is associated with nonverbal ability while the chromatic contrast discrimination task is not. This result impacts upon the appropriateness of the clinical use of the FM100. The results from this study suggest that the use of an appropriate psychophysical task, which has fewer task demands and is of equal difficulty across all ages will give a more accurate measure of color discrimination and ultimately visual function in participants.

Acknowledgments

The authors thank the teachers and students at Tasker House, Newcastle College, Barndale House School, South Gosforth First School, Bamburgh School, and Collingwood School, for their help and participation. They also thank undergraduate students Charlotte Reay, Max O'Collins, and Vikki Ingram for research assistance, the two anonymous reviewers for their very helpful comments on an earlier draft of this manuscript, and Deborah Riby and Ann Le Couteur for their general supervisory support.

Supported by grants from the Estate of David Murray Garside (MBC; Newcastle Upon Tyne, UK), the Newcastle Vision Fund (MBC; Newcastle Upon Tyne, UK), and the Engineering and Physical Sciences Research Council (EP/H022325/1; Swindon, UK).

Disclosure: **M.B. Cranwell**, None; **B. Pearce**, None; **C. Love-ridge**, None; **A.C. Hurlbert**, None

References

1. Treisman AM, Gelade G. A feature-integration theory of attention. *Cogn Psychol*. 1980;12:97-136.
2. Tanaka J, Weiskopf D, Williams P. The role of color in high-level vision. *Trends Cogn Sci*. 2001;5:211-215.
3. Whitehead RD, Re D, Xiao D, Ozakinci G, Perrett DI. You are what you eat: within-subject increases in fruit and vegetable consumption confer beneficial skin-color changes. *PLoS One*. 2012;7:e32988.
4. Roberson D, Davies I, Davidoff J. Color categories are not universal: replications and new evidence from a stone-age culture. *J Exp Psychol Gen*. 2000;129:369-398.
5. Palmer SE, Schloss KB. An ecological valence theory of human color preference. *Proc Natl Acad Sci U S A*. 2010;107:8877-8882.
6. Ling Y, Hurlbert A. Role of color memory in successive color constancy. *J Opt Soc Am A*. 2008;25:1215-1226.
7. Franklin A, Sowden P, Burley R, Notman L, Alder E. Color perception in children with autism. *J Autism Dev Disord*. 2008;38:1837-1847.
8. Franklin A, Sowden P, Notman L, et al. Reduced chromatic discrimination in children with autism spectrum disorders. *Dev Sci*. 2010;13:188-200.
9. Hurlbert A, Loveridge C, Ling Y, Kourkoulou A, Leekam S. Color discrimination and preference in autism spectrum disorder. *J Vis*. 2011;11(3):429.
10. Banaschewski T, Ruppert S, Tannock R, et al. Colour perception in ADHD. *J Child Psychol Psych*. 2006;47:568-572.

11. Kim S, Chen S, Tannock R. Visual function and color vision in adults with Attention-Deficit/Hyperactivity Disorder. *J Optometry*. 2014;7:22-36.
12. Farran EK, Cranwell MB, Alvarez J, Franklin A. Colour discrimination and categorisation in Williams syndrome. *Res Dev Disabil*. 2013;34:3352-3360.
13. Galton F. *Inquiries Into Human Faculty and its Development*. London: Dent; 1883.
14. Spearman C. "General intelligence," objectively determined and measured. *Am J Psychol*. 1904;15:201-292.
15. Evans RB, Deary IJ. Sensory discrimination and intelligence: postmortem or resurrection? *Am J Psychol*. 1994;107:95-115.
16. Acton GS, Schroeder DH. Sensory discrimination as related to general intelligence. *Intelligence*. 2001;29:263-271.
17. Deary IJ. Intelligence and auditory discrimination: separating processing speed and fidelity of stimulus representation. *Intelligence*. 1994;18:189-213.
18. Li S-C, Jordanova M, Lindenberger U. From good senses to good sense: a link between tactile information processing and intelligence. *Intelligence*. 1998;26:99-122.
19. Melnick MD, Harrison BR, Park S, Bennetto L, Tadin D. A strong interactive link between sensory discriminations and intelligence. *Curr Biol*. 2013;23:1013-1017.
20. Farnsworth D. *The Farnsworth-Munsell 100-Hue Test for the Examination of Color Discrimination: Manual*. Baltimore: Macbeth; 1957.
21. Dain SJ. Clinical colour vision tests. *Clin Exp Opt*. 2004;87:276-293.
22. Kinnear PR, Sahraie A. New Farnsworth-Munsell 100 hue test norms of normal observers for each year of age 5-22 and for age decades 30-70. *Br J Ophthalmol*. 2002;86:1408-1411.
23. Gunther KL, Neitz JAY, Neitz M. A novel mutation in the short-wavelength-sensitive cone pigment gene associated with a tritan color vision defect. *Vis Neurosci*. 2006;23:403-409.
24. Heywood CA, Gadotti A, Cowey A. Cortical area V4 and its role in the perception of color. *J Neurosci*. 1992;12:4056-4065.
25. Ménage MJ, Papakostopoulos D, Dean Hart JC, Papakostopoulos S, Gogolitsyn Y. The Farnsworth-Munsell 100 hue test in the first episode of demyelinating optic neuritis. *Br J Ophthalmol*. 1993;77:68-74.
26. Victor JD, Maiese K, Shapley R, Sidtis J, Gazzaniga MS. Acquired central dyschromatopsia: analysis of a case with preservation of color discrimination. *Clin Vision Sci*. 4:183-196.
27. Knoblauch K, Saunders F, Kusuda M, et al. Age and illuminance effects in the Farnsworth-Munsell 100-hue test. *Appl Opt*. 1987;26:1441-1448.
28. Beirne RO, McIlreavy L, Zlatkova MB. The effect of age-related lens yellowing on Farnsworth-Munsell 100 hue error score. *Ophthalmic Physiol Opt*. 2008;28:448-456.
29. Birch J, Barbur JL, Harlow AJ. New method based on random luminance masking for measuring isochromatic zones using high resolution colour displays. *Ophthalmic Physiol Opt*. 1992;12:133-136.
30. Regan BC, Reffin JP, Mollon JD. Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. *Vision Res*. 1994;34:1279-1299.
31. Paramei GV, Oakley B. Variation of color discrimination across the life span. *J Optical Soc Am A*. 2014;31:A375-A384.
32. Barbur JL, Rodriguez-Carmona M. Colour vision changes in normal aging. In: Elliot AJ, Fairchild MD, eds. *Handbook of Color Psychology*. Cambridge, UK: Cambridge University Press. In press.
33. Knoblauch K, Vital-Durand F, Barbur JL. Variation of chromatic sensitivity across the life span. *Vision Res*. 2001;41:23-36.
34. Birch J. *Diagnosis of Defective Colour Vision*. 2nd ed. Oxford, UK: Butterworth Heinemann; 2001.
35. Karpf RJ, Goss AE, Small MY. Naming, selection, and ordering of color ("hue") by young children. *J Gen Psychol*. 1974;90:297-314.
36. Dain SJ, Ling BY. Cognitive abilities of children on a gray seriation test. *Optom Vis Sci*. 2009;86:E701-E707.
37. Schroeder A, Kreutz M, Meyer M, Erb C. Influence of learning effects on the results of the cap-sorting test Roth 28-hue (E) desaturated: influence of learning effects on colour-arrangement tests. *Color Res Appl*. 2007;32:16-21.
38. Foote KG, Neitz M, Neitz J. Comparison of the Richmond HRR 4th edition and Farnsworth-Munsell 100 Hue Test for quantitative assessment of tritan color deficiencies. *J Opt Soc Am A*. 2014;31:A186-A188.
39. Wechsler D. *Wechsler Abbreviated Scale of Intelligence*. San Antonio: Psychological Corporation; 1999.
40. Wechsler D. *Wechsler Intelligence Scale for Children - Fourth Edition*. San Antonio: Harcourt Assessment, Inc.; 2003.
41. Neitz J, Summerfelt P, Neitz M. *The Neitz Test of Color Vision*. Los Angeles: Western Psychological Service; 2001.
42. Brainard DH. The Psychophysics Toolbox. *Spat Vis*. 1997;10:433-436.
43. Wolf K. *The Modulation of Simultaneous Chromatic Contrast*. Institute of Neuroscience. Newcastle Upon Tyne: University of Newcastle Upon Tyne; 2011.
44. Eskew RT, McLellan JS, Giulianini F. Chromatic detection and discrimination. In: Gegenfurtner KR, Sharpe LT, eds. *Color Vision: From Genes to Perception*. Cambridge, UK: University Press; 1999:345-368.
45. Kramer JH, Ellenberg L, Leonard J, Share LJ. Developmental sex differences in global-local perceptual bias. *Neuropsychology*. 1996;10:402-407.
46. Plaisted K, Swettenham J, Rees L. Children with autism show local precedence in a divided attention task and global precedence in a selective attention task. *J Child Psychol Psychiatry*. 1999;40:733-742.
47. Rinehart NJ, Bradshaw JL, Moss SA, Brereton AV, Tonge BJ. A deficit in shifting attention present in high-functioning autism but not Asperger's disorder. *Autism*. 2001;5:67-80.
48. Goulart PRK, Bandeira ML, Tsubota D, Oiwa NN, Costa ME, Ventura DE. A computer-controlled color vision test for children based on the Cambridge Colour Test. *Vis Neurosci*. 2008;25:445-450.
49. Barbur JL, Harlow AJ, Plant GT. Insights into the different exploits of colour in the visual cortex. *Proc Biol Sci*. 1994;258:327-334.
50. Parrish EE, Giaschi DE, Boden C, Dougherty R. The maturation of form and motion perception in school age children. *Vision Res*. 2005;45:827-837.
51. McAvinue LP, Habekost T, Johnson KA, et al. Sustained attention, attentional selectivity, and attentional capacity across the lifespan. *Atten Percept Psychophys*. 2012;74:1570-1582.