

Chapter 12

From Subtypes to Taxons: Identifying Distinctive Profiles of Reading Development in Children



Adrian P. Burgess, Caroline Witton, Laura Shapiro, and Joel B. Talcott

Abstract The longstanding debate between dimensional and categorical approaches to reading difficulties has recently been rekindled by new empirical evidence and developments in theory. At the heart of the categorical perspective is the tenet that dyslexia is a taxon, a grouping of cases that can account for both intra-group similarities and inter-group differences. As developmental dyslexia is characterized by a diverse constellation of symptoms with multiple underlying risk and protective factors, the key question in dyslexia research has shifted from “What is dyslexia?” to “How many taxons or subtypes of dyslexia are there?” The primary objective of this chapter is to consider methods that can be used to objectively define these groupings, starting with the current practice of defining subtypes of readers using normative scores with pragmatically dened cut-offs, the “Quadrant Analysis” approach, and progressing towards more theoretically sound and statistically rigorous procedures. We review and test several candidate approaches that can be readily adapted to realistic conditions that are problematic for Quadrant Analysis. Specifically we propose a method that can be used to identify subgroups in the bivariate case when the two indicator variables are correlated. We conclude by evaluating the strengths and weaknesses of this and other methods and include implications for their future application toward identifying and validating putative dyslexia taxons.

Keywords Dyslexia · Subtypes · Orthography · Phonology · Taxon · Development · Reading · Statistical modelling

A. P. Burgess (✉) · C. Witton · L. Shapiro · J. B. Talcott
Aston Brain Centre, Aston University, Birmingham, UK
e-mail: a.burgess@aston.ac.uk; c.witton@aston.ac.uk; l.shapiro@aston.ac.uk;
j.b.talcott@aston.ac.uk

12.1 Introduction: Categories or Dimensions?

That some children are poorer readers than others is beyond dispute and that some are poorer than might be expected given their overall cognitive ability is also certain. However, it remains unclear whether poor readers are simply statistical outliers (as is inevitable in any norm-based measurement system) or whether they form a subgroup that differs from typical readers in certain key characteristics (see e.g., S. E. Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch 1992). This debate, between dimensional and categorical approaches to children with reading difficulties, is longstanding but has recently been revived by the decision to replace the term “dyslexia” in DSM-5 with “specific learning disability (reading)” (*Diagnostic and statistical manual of mental disorders: DSM-5*, American Psychiatric Association 2013) and by resurgent claims that dyslexia is an unscientific construct that does not really exist (Elliott & Grigorenko 2014). This is no arcane debate because the true nature of specific reading difficulties has powerful implications for the diagnosis, intervention and service provision for those affected and their families.

The heart of the categorical perspective on reading difficulties is that dyslexia forms a taxon. That is, dyslexia constitutes a grouping of cases that share underlying commonalities that not only account for the similarities between group members but also explain how and why group members differ from non-group members. Sex is an example of a taxon; males and females share many common features but differ in certain fundamental characteristics that justify considering sex as a categorical construct. More formally, a taxon is a fundamental, objective, non-arbitrary and reasonably enduring latent structure (Ruscio & Ruscio 2004). To justify dyslexia as a taxon therefore, children with reading difficulties should either show some characteristics that are qualitatively distinctive, or the distribution of their latent abilities should be discontinuous from those of typical readers.

In fact, given that developmental dyslexia is characterized by a diverse constellation of symptoms with multiple underlying risk and protective factors (Pennington 2006) the question often raised is not “*Does dyslexia constitute a taxon*” but “*How many taxons or subtypes of dyslexia are there?*” (Pennington 2006; Peterson, Pennington, Olson, & Wadsworth 2014) and this is currently an active area of research. To date, this pursuit has been most successful in identifying individuals with presumed dissociations between cognitive skills closely linked to reading achievement, for example phonological and orthographic skills consistent with that predicted by the dual route model (Castles & Coltheart 1993), or separable dimensions of phonological decoding and articulatory naming speed (i.e., rapid automatized naming (RAN)) as consistent with the double deficit hypothesis (Bowers & Wolf 1993). Both models predict the occurrence of discrete subtypes of individuals, with relatively isolated deficits in a single component process and comparatively normal functioning in the other and each approach has reported subtypes of dyslexia that are largely consistent with its own theoretical perspective.

The definition of subtypes of dyslexia primarily derives from normative performance and membership is assigned to those who score below a specified level

of performance on one or more theoretically relevant cognitive tasks. Defined in this way, any continuous bivariate distribution will necessarily divide the whole sample into quadrants: those who score above the threshold on both dimensions, those who score below the threshold on both dimensions and those who score above the threshold on one dimension and below the threshold on the other. The choice of threshold is at least semi-arbitrary and may be determined pragmatically by non-theoretical considerations. For example, the threshold might be set to ensure that there are sufficient cases in each group to allow statistical analysis or to equate to a round number of stanines, z-scores or percentiles. Of course, the arbitrary nature of such thresholds precludes these subtypes from being considered as taxons.

It is not that there is anything fundamentally wrong with using cut-off scores with continuous variables (categorization of continuous data is commonplace and often useful) but we should be clear that is what we are doing, if that is what we are doing. If reading ability is dimensional, then we should be clear that the subtypes identified are not fundamental and are comparable to groups like the “tall” or the “rich” which can be defined in many different ways. If, on the other hand, reading profiles are taxonic, we should be explicit about how the subtypes differ and set cut-off scores at a level that optimally separate the groups.

The primary objective of this chapter is to consider ways in which we can progress from defining subtypes of readers using normative scores with pragmatically defined cut-offs and move towards a more rigorous approach to identifying reading taxons, if they exist. In the first part, we address the use of normative scores with cut-offs to identify subtypes of readers, a method we refer to as “Quadrant Analysis”. Specifically, we show how to estimate the proportion of children in each subtype and argue that deviations between the observed and expected numbers of children in each quadrant might provide useful information about where cut-off scores should be set. In the second part, we develop this idea and propose a method that makes the choice of thresholds less arbitrary and illustrate its use with data from a previously published study (Talcott et al. 2002). Finally, we briefly consider alternative approaches to answering the “*dimension or category?*” question.

12.2 Using Normative Scores to Define Subtypes: Quadrant Analysis

The identification of one or more subtypes of children who show qualitatively different profiles of reading ability from typical developing children has grown into an area of considerable research interest, if little consensus. There is, currently, no agreed definition or characterization of these subtypes and this absence has encouraged the emergence of multiple arbitrary and ad-hoc definitions of subgroups of children based on performance on one or more measures of reading ability (Boder 1973; Castles & Coltheart 1993) or other cognitive dimensions (Bosse, Tainturier, & Valdois 2007; Bowers & Wolf 1993). For example, there is evidence to suggest

that the ability to read phonetically regular non-words such as “*tegwog*” is at least partially dissociated from the ability to read phonetically irregular real words such as “*yacht*” in learner readers. To develop this example, one can define subgroups from these two measures (non-words vs. exception words) by defining some criteria for good and bad performance on each of these tests (Sprenger-Charolles, Siegel, Jiménez, & Ziegler 2011). This defines four groups: (i) those who are good at reading both non-words and exception words, (ii) those who are bad at reading both types, (iii) those who are good at reading non-words but poor at exception words and (iv) those who are good at reading exception words but poor at reading non-words. The term “Quadrant Analysis” derives from the fact that this approach inevitably creates four groups, but the groups are not normally equal in size and the cut-offs may be used to define two groups (one quadrant vs. the other three). It should be obvious that procedures of this type will inevitably identify a subset of children regardless of whether dyslexia exists or not and so is of no use in identifying taxons.

The choice of cognitive skills to measure (e.g., non-word reading vs exception word reading) is at least driven by theoretical considerations. In contrast, the choice of threshold of determining “good” and “bad” performance is almost completely open. The threshold is typically defined such that any score less than a certain number of standard deviations below the mean is considered to indicate poor performance. In the absence of any good evidence as to how such a cut-off should be defined, it is usual to select a non-arbitrary seeming number of standard deviations below the mean (1/2, 1, 1 1/2), a whole number of stanines (1, 2, 3 or 4) or a round number percentile (10%, 20%, etc.). Such cut-off scores are chosen to appear principled and may allay the suspicion that the definition is opportunistic (imagine your response to reading that a definition of 0.6745 standard deviations below the mean was used¹) but these cut-offs are essentially arbitrary and are at least as much determined by convenience and an affection for round numbers as they are by empirical evidence.

More sophisticated, multivariate methods, such as identifying poor readers based on a discrepancy between their actual reading ability and their predicted reading ability (for example, that based on a regression of reading achievement on cognitive ability) may have some advantages but every method that uses normative scores and performance cut-offs to define subtypes of dyslexia, shares the same limitations. First, they will always identify subtypes of poor readers whether or not any distinctive taxon, such as dyslexia, exists. Second, they are fundamentally arbitrary and are not based on the distinctive characteristics of the taxon (assuming the taxon exists at all).

The absence of any objectively defined criterion or cut-off scores means that individual researchers are at liberty to define reading subtypes in whatever way suits them and it is a liberty freely exercised. Worse still, they can change their criteria from study to study at their own convenience. Unsurprisingly, the consequence of this approach is that reading subtypes are inconsistently defined in the scientific

¹Actually 0.6745 is the 25th percentile so perhaps not as arbitrary as it first appears.

literature and the justification for the cut-off scores used is not explained. It seems likely that many of the inconsistencies that have been reported, and the controversies they inspire, result at least in part from such failures of definition. None of this should be a cause for surprise because Quadrant Analysis is not designed to identify subtypes, let alone taxons. Indeed, the use of Quadrant Analysis to define reading subtypes can only be defended because the current state of knowledge makes it difficult to know what else to do and, although there are alternative approaches (e.g., Cluster Analysis, Finite Mixture Modeling, and Taxometric Methods – see Sect. 12.5), these too have their problems. Nevertheless, as Quadrant Analysis is widely used, it is worth exploring how the method might be improved.

In one wished to argue that these Quadrant Analysis “subtypes” are taxonic, it would be necessary to provide some principled justification for the choice of cut-off. For example, the subtypes should show some characteristic that would not be predicted by assuming the observed scores came from a continuous bivariate distribution. One such characteristic might be that the proportion of individuals in each quadrant differs from what would be expected if the data derived from a continuous bivariate distribution and this is the case we consider here. This is of interest because if learner readers consist of one or more subtypes, then the scores will not be normally distributed because each of the subtypes will have their own mean, standard deviation and covariance. In extreme cases, this will result in a bimodal distribution of scores although this is relatively uncommon in human performance data. More generally, a unimodal distribution will be seen but the number of individuals in each quadrant (defined by cut-offs determined by the mean and standard deviation of the whole sample) will differ from what would be seen if the data were normally distributed. This means that the proportion of individuals observed in each quadrant, compared to the expected number may offer some information on whether a subtype exists or not and where the cut-offs should be positioned. To this end, in the following section we show how to estimate the observed and expected proportions of cases in each quadrant.

12.3 How Many Individuals in Each Quadrant?

Let’s consider the example of reading non-words and exception words, and consider the case where we wish to know the proportion of individuals that score below a given cut-off score on these two measures and compare that to the proportion of cases that would be expected if the data came from bivariate normal distribution with no subtypes. If the scores on the two tests are independent (i.e., not correlated) then the problem is quite straightforward. To start, let’s use the mean as the cut-off score (i.e., 0 standard deviations below the mean). If the data are normally distributed, we would expect to find half the sample scoring below the mean on each of our tests and half above. That is, the probability of scoring below the mean on either test is 0.5.

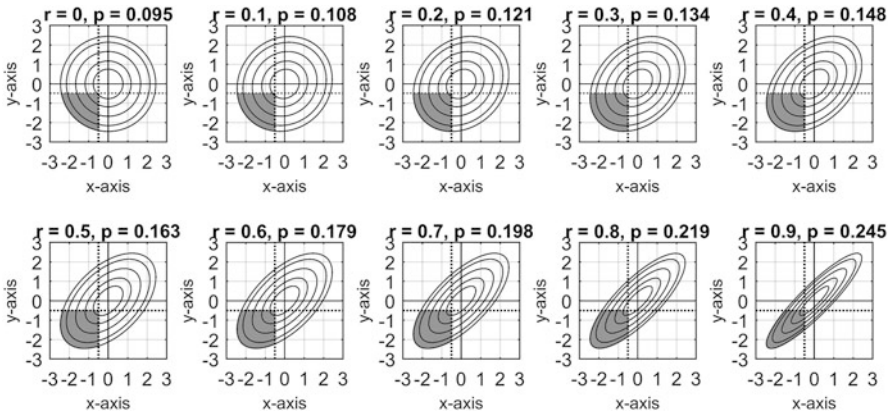


Fig. 12.1 Showing the PDF of the bivariate normal distribution with correlations from 0.0 to 0.9 and standard deviation = 1. The rings indicate the regions containing (from inner to outer) 25%, 50%, 90% and 95% of the population. The shaded area indicates the proportion of the population that is more than 1/2 standard deviation below the mean on both measures which increases from 9.5% when $r = 0$ to 24.5% when $r = 0.9$

If the test scores are uncorrelated then, by the multiplication rule,² the proportion of individuals scoring below the mean on both tests will be $0.5 \times 0.5 = 0.25$. For other cut-offs, we need to know the probability density function (PDF) of the normal distribution which is given by:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-(x - \mu)^2 / (2\sigma^2)\right] \tag{12.1}$$

Where μ is the population mean and σ^2 is the population variance. In order to find the proportion of the population expected to score below any given cut-off, x_{cut} , we need to calculate the cumulative probability from $-\infty$ to x_{cut} , which, can be estimated through numerical integration of Eq. 12.1 and is readily available in Tables. For example, if we use a cut-off score of 1/2 standard deviation below the mean, we know from the cumulative PDF of the normal distribution that approximately 31% of individuals fall below this level. It follows that about 9.5% of individuals would be expected to score 1/2 standard deviation or more below the mean on two tests ($0.31 \times 0.31 = 0.095$).

If the scores on the tests are correlated (as they usually will be), the situation becomes more complicated and the proportion of participants expected to score below any given cut-off of two tests will increase with the correlation (Fig. 12.1). For example, the proportion of individuals scoring 1/2 standard deviation or more below the mean on both tests ranges from 0.095 when the correlation is 0 to 0.31,

²The multiplication rule, $p(A \cap B) = p(A) \cdot p(B/A)$ which is $p(B/A) = p(B)$ when A and B are independent.

when the correlation is +1 (Fig. 12.1). To estimate this proportion for any given correlation, we need to know the PDF of the bivariate normal distribution which is given by:

$$P(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left[\frac{\left(\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right)}{2(1-\rho^2)} \right] \quad (12.2)$$

Where ρ is the correlation between x and y . To estimate the proportion of the population expected to score below a specified cut-off, x_{cut} , on test x and below a specified cut-off, y_{cut} , on test y , we simply have to estimate the cumulative probability of Eq. 12.2 $-\infty$ to x_{cut} and $-\infty$ to y_{cut} . For the convenience of the reader Table 12.2 was produced which shows the proportion of the population expected to score below a given cut-off score (ranging from 0 to -2 standard deviations below the mean) on two tests for correlations between 0 and 0.9 and a summary of the same data is represented graphically in Fig. 12.2. As can be seen from Fig. 12.2, the proportion of participants scoring below both cut-offs tends towards zero as the correlation, ρ , approaches -1 and tends towards the univariate marginal probability defined by the cut-off as ρ , approaches $+1$.

It would also be useful to have confidence intervals for these proportions for use with empirical data and this can be estimated assuming the binomial distribution. It follows that if p is the proportion of the population that will score below a specified cut-off, x_{cut} , on test x **and** below a specified cut-off, y_{cut} , on test y , then the expected value of sampling from the population is np where n is the sample size and the variance of the estimate will be $np(1-p)$. From these values, it is straightforward to estimate confidence intervals for and desired combination of cut-off score, correlation and sample size.

To illustrate the process, we show how to estimate the proportion of participants in each quadrant and illustrate the method using data from the Oxford Primary School Study (Talcott et al. 2002). This sample comprised 353 children (183 girls, 170 boys) between the ages of 83 and 150 months (mean 112.8, s.d. 14.9). All of the children attended mainstream primary schools within the local education authority. Children who did not have English as a first language were not included in the study, but no other selection criterion was applied.

Whichever approach is used to identify latent taxonic structure, it is necessary to select appropriate indicator variables that measure the construct in question. In the case of reading, a very large number of indicators have been used that purport to discriminate between typical and dyslexic readers but, influenced by dual-route models of reading (Coltheart, Rastle, Perry, Langdon, & Ziegler 2001), we elected to use two distinct measures of reading ability. Specifically, we chose non-word reading as the putative measure of phonological processing and exception word reading as the putative measure of orthographic processing. In languages with opaque orthographies as English, the inconsistency of the mapping between

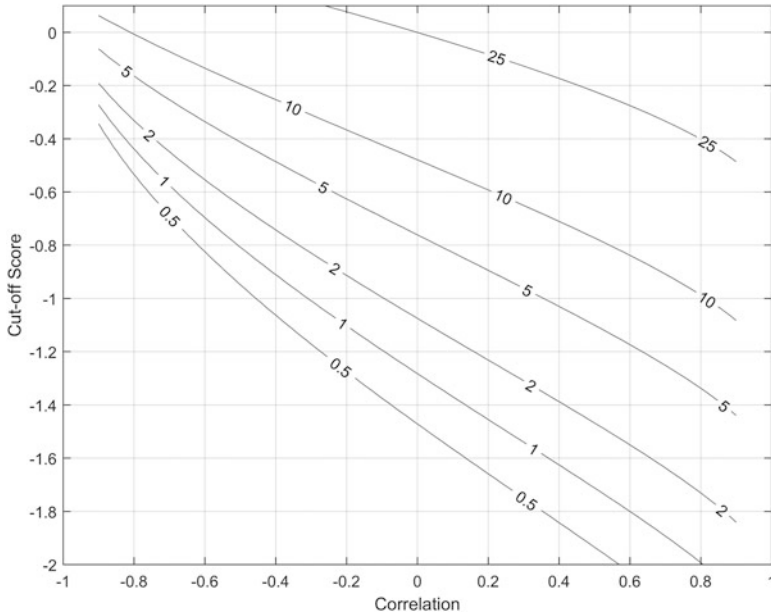


Fig. 12.2 Contour plot of the proportion of participants expected to score below a given cut-off score both of two tests by the correlation between the tests. The cut-off scores are defined as the number of standard deviations below the mean. Labels on contours are percentiles

letters and sounds poses a difficult task for the beginning reader, and ultimately requires a development of a reading system that is flexible, with lexical access facilitated by both the phonological and orthographic characteristics of words (Coltheart 1978). Although the strongest determinant of reading aptitude in typically developing children, and of reading impairments, is the competency to which phonological decoding skills are acquired and employed (Coltheart 1978). Lexical access in simple reading tasks (Rack, Snowling, & Olson 1992; Wagner & Torgesen 1987), impairments of other reading sub-skills, such as in orthographic coding, also explain variance in literacy skill in some individuals with developmental dyslexia (Badian 2005; Castles & Coltheart 1993). There is also strong evidence that phonological and orthographic impairments contribute independently to the heritable and presumed genetic component of risk for specific reading difficulties (Castles, Datta, Gayan, & Olson 1999). Evidence that subtypes of dyslexia based on dual route models may have different developmental trajectories, with implications for assessment and intervention (Manis, Seidenberg, Doi, McBride-Chang, & Petersen 1996; Talcott, Witton, & Stein 2013).

Children were assessed on a wide range of measures (Table 12.1) but the indicator variables we chose for the following examples was the Castles and Coltheart Reading Test (1993) which provides reading scores for exception words and non-words in the range 0–30.

Table 12.1 Cognitive Performance on selected tests for the Oxford Primary School Study (Talcott et al. 2002)

Cognitive test	Mean (s.d.), skew
BAS ^a verbal IQ	58.8 (9.0), -0.28
BAS matrices	56.1 (8.8), -0.20
BAS reading	54.8 (10.9), -0.16
BAS spelling	52.4 (8.9), 0.02
C&C ^b regular word reading	24.8 (6.8), -1.70
C&C non-word reading	19.4 (8.0), -0.65
C&C exception word reading	16.4 (6.1), -0.71

^aBAS = British Ability Scales (Elliot, Murray, & Pearson 1983)

^bC&C = Castles and Coltheart Reading Test

The mean score for non-words was 19.4 (s.d. = 8.0) and for Exception words was 16.4 (s.d. = 6.1) and the correlation between the two was 0.76. A scatterplot for the observed scores from this sample is shown in Fig. 12.3a along with the estimated marginal PDF, estimated using kernel smoothing, for non-words (Fig. 12.3b) and exception-words (Fig. 12.3c). Each of the marginal PDFs is shown with a normal PDF with the same mean and standard deviation as the observed PDF. PDFs were obtained using Kernel Density Estimation (also known as the Parzen-Rosenblatt window method) which is a non-parametric method for estimating the PDF of a random variable that provides more robust and reliable estimates of the true PDF than traditional histogram methods.

Seventy children out of 353 obtained a score of less than 1/2 standard deviations below the mean on both tests which is 19.8% of the whole sample. The question we address here is whether this proportion is higher than would be expected if the data followed a bivariate normal distribution? To begin, it is worth noting that neither marginal distribution appears to be normally distributed and this suspicion is supported by the Shapiro-Wilk test which shows that both distributions deviate significantly from normal (Non-words: $W = 0.915$, $p < 0.01$; Exception Words: $W = 0.954$, $p < 0.01$) and that, in consequence, the data cannot be bivariate normal. For both tests, the data were truncated by the range of possible scores and there is evidence for a ceiling effect on nonword word reading. In addition, the scores can only take integer values and are not truly continuous as would be expected in a normal distribution. Nevertheless, the PDFs of the marginal distributions are not so abnormal that the idea of using conventional parametric statistical analysis with them (e.g., t-tests, analysis of variance (ANOVA), etc.) would cause much concern (Fig. 12.4).

We can compare the proportion of participants in the observed sample who score below the cut-off of $-1/2$ standard deviations on both tests with the proportions expected from a bivariate normal distribution using the data in Table 12.2. The closest entries to a cut-off of $-1/2$ and a correlation of 0.76 are 0.198 and 0.219 for $r = 0.7$ and 0.8 respectively. By linear interpolation this gives 0.211 for $r = 0.76$ which is close to the observed proportion of 0.198 but it remains to be determined whether a difference of this size is likely to be real or simply due to chance

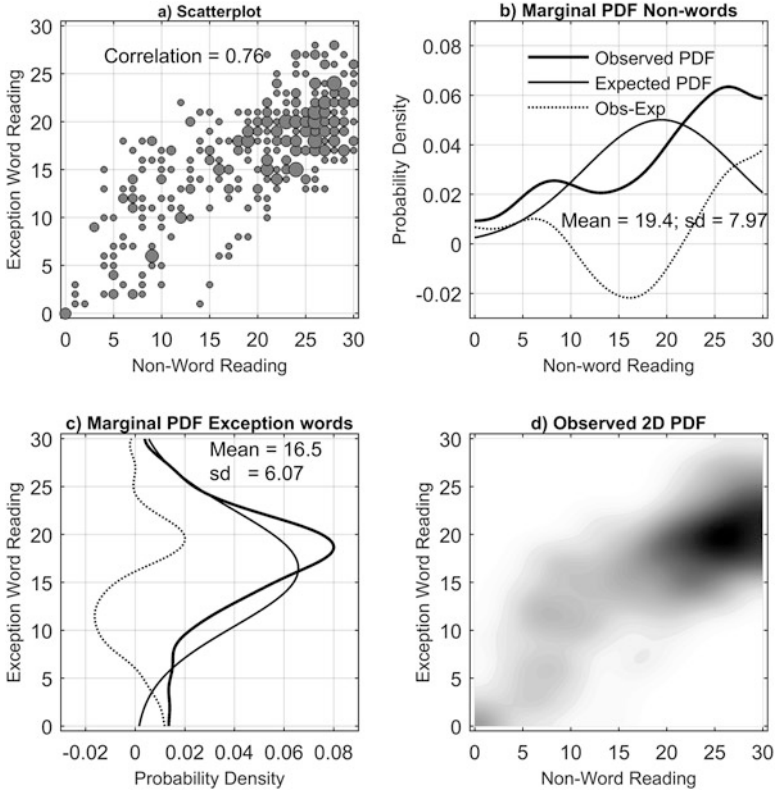


Fig. 12.3 Oxford Primary School Study cohort showing (a) the scatterplots of non-word reading by exception word reading for, (b) the marginal distribution of non-word reading (thick line) with the best fitting normal distribution (thin line), (c) the marginal distribution of exception-word reading (thick line) with the best fitting normal distribution (thin line) and (d) the 2-dimensional PDF of non-word reading by exception word reading (darker colours indicate higher density)

variation. Using binomial theory, and assuming we were sampling from a bivariate normal distribution, the mean expected number of people we should expect to see scoring below cut-off on both tests is np i.e. $353 \times 0.211 = 74.5$ and the variance is $np(1 - p)$ which is $353 \times 0.211 \times (1 - 0.211) = 58.8$ giving a standard deviation of $\sqrt{58.8} = 7.7$. To get the 95% confidence intervals we calculated the mean score ± 2 standard deviations which gives $74.5 \pm 2 \times 7.7$ giving 59–90 (rounding down and up to the nearest integer respectively). As the observed number of cases ($n = 70$) was within this interval, we can conclude that the number of cases observed is within the bounds that would be expected if we were sampling from a bivariate normal distribution.

The illustration here is used to examine individuals who scored below the cut-off on both tests but the same ideas can be applied to each of the other quadrants if desired. We know from the cumulative probability function of the

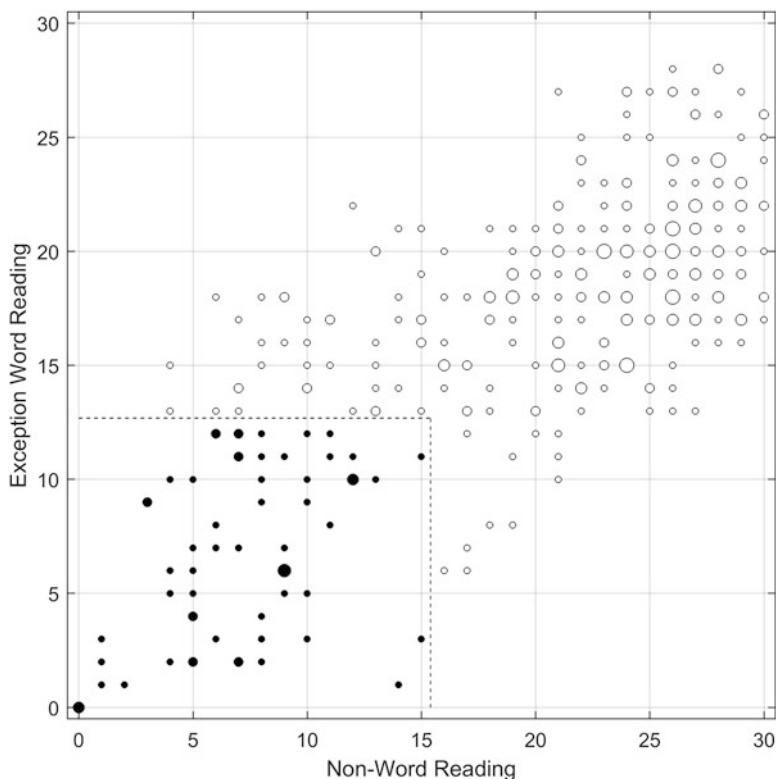


Fig. 12.4 Oxford Primary School Study cohort showing the classification of cases by the Quadrant method using cut-off scores based on the point of maximum difference between the observed and expected PDFs. Dotted lines indicate the borders of the lower-left hand quadrant

normal distribution that the probability of a score less than $-1/2$ standard deviations below the mean is 0.309. From the example above, we know that the proportion of participants who would be expected to score less than $-1/2$ on both tests was 0.211 so it follows that those who scored below $-1/2$ on Test 1 and above $-1/2$ on Test 2 is $0.309 - 0.211 = 0.098$. By symmetry, the proportion who scored above $-1/2$ on Test 1 and below $-1/2$ on Test 2 is the same, 0.098. Knowing the proportion of cases in three of the four quadrants, the fourth quadrant is not hard to find.

The finding that the proportion of participants in each quadrant is within the bounds of what would be expected with a bivariate normal distribution with the same correlation is consistent with data from other studies. Using the same procedure with published data on dyslexia subtypes (Castles & Coltheart 1993; Genard, Mousty, Content, Alegria, Leybaert, & Morais 1998; Jimenez, Rodriguez, & Ramirez 2009; Manis et al. 1996; Sprenger-Charolles, Colé, Lacert, & Serniclaes 2000; Ziegler 2008) suggests that the frequencies observed in each group in empirical studies often fail to differ significantly from what would be expected if the data had been

Table 12.2 Showing the proportion of cases expected to score below the z-score cut-off on both tests for given levels of correlation between the tests assuming a bivariate normal distribution with correlations between 0 and 0.9. Probabilities derived from Eq. 12.2 were estimated by numerical integration in MatLab (R2015a) using function integral2.m over the range -4 to x_{cut} and -4 to y_{cut} with $dx = dy = 0.01$

z-score cut-off	Correlation between tests									
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.00	0.250	0.266	0.282	0.298	0.315	0.333	0.352	0.373	0.398	0.428
-0.10	0.212	0.228	0.243	0.260	0.277	0.294	0.313	0.334	0.358	0.389
-0.20	0.177	0.192	0.208	0.224	0.240	0.258	0.276	0.297	0.320	0.350
-0.30	0.146	0.161	0.175	0.191	0.207	0.223	0.241	0.261	0.284	0.313
-0.40	0.119	0.132	0.146	0.161	0.176	0.192	0.209	0.228	0.250	0.278
-0.50	0.095	0.108	0.121	0.134	0.148	0.163	0.179	0.198	0.219	0.245
-0.60	0.075	0.087	0.098	0.111	0.124	0.138	0.153	0.170	0.189	0.214
-0.70	0.059	0.069	0.079	0.090	0.102	0.115	0.129	0.144	0.162	0.186
-0.80	0.045	0.054	0.063	0.073	0.083	0.095	0.107	0.121	0.138	0.160
-0.90	0.034	0.041	0.049	0.058	0.067	0.077	0.089	0.101	0.117	0.136
-1.00	0.025	0.031	0.038	0.045	0.054	0.062	0.072	0.084	0.098	0.115
-1.10	0.018	0.023	0.029	0.035	0.042	0.050	0.059	0.069	0.081	0.097
-1.20	0.013	0.017	0.022	0.027	0.033	0.039	0.047	0.056	0.066	0.081
-1.30	0.009	0.013	0.016	0.020	0.025	0.031	0.037	0.045	0.054	0.066
-1.40	0.007	0.009	0.012	0.015	0.019	0.024	0.029	0.036	0.044	0.054
-1.50	0.004	0.006	0.009	0.011	0.015	0.018	0.023	0.028	0.035	0.044
-1.60	0.003	0.004	0.006	0.008	0.011	0.014	0.017	0.022	0.028	0.035
-1.70	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.017	0.022	0.028
-1.80	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.017	0.022
-1.90	0.001	0.001	0.002	0.003	0.004	0.006	0.007	0.010	0.013	0.017
-2.00	0.001	0.001	0.001	0.002	0.003	0.004	0.005	0.007	0.010	0.013

drawn from a bivariate normal distribution and, by definition, a bivariate normal distribution suggests a single population with no subtypes or taxons. This provides a challenge to the use of this simple dissociation logic to define subgroups of dyslexia in this context and suggests that stronger evidence is required in order to reify the existence of distinct diagnostic entities from such data.

12.4 An Improved Method for Choosing Cut-Off Scores

In the case of the Oxford Primary School Study (Talcott et al. 2002), the proportion of cases in each quadrant is very close to what might be expected had the data been drawn from a bivariate normal distribution but that does not completely exclude the possibility that subtypes exist. Indeed, the distribution of performance on the non-word reading test appears to show a bimodal distribution (Fig. 12.3b) and the two-dimensional-PDF (Fig. 12.3d) shows two distinct regions of high density. This suggests that if the cut-off scores had been better placed, they might have provided a subdivision of the sample into the clusters that visual inspection seems to suggest exist. It is obvious that the optimal location for the cut-off scores is where the difference between the observed and expected numbers of cases is greatest and this can be found by extending the logic of the above example in which the difference between the observed and expected number of cases are calculated with cut-offs of $-1/2$ standard deviations to all possible cut-offs. To do this, it is only necessary to estimate the observed and expected PDFs.

By assuming a bivariate normal distribution with the same means, standard deviations and correlation as the observed data, it is straightforward to estimate the expected PDF and this is shown in Fig. 12.3c. Estimation of the observed PDF can be done in a number of different ways but we used a 2D kernel estimator (Botev, Grotowski, & Kroese 2010) using their MatLab function “kde2d.m”, available from Botev (2015). The estimated observed PDF is shown in Fig. 12.3e. Figure 12.3f shows the difference between the observed and expected PDFs. The difference between observed and expected values represents the greatest difference and shows a minimum at non-words = 15.6 and exception Words = 12.9, which corresponds to z-scores of -0.473 and -0.583 respectively. Assuming bivariate normality with $r=0.76$, the probability of scoring below both cut-off scores, estimated, using numerical integration, was 0.201. This gives the expected number of cases as 71 ± 15 (i.e., 56–86). The first group was defined as those individuals who scored below the cut-off score on both measures and this amounted to 63 individuals making up 17.8% of the sample which did not differ significantly from what was expected. The second group consisted of all those individuals not in the first group and, by definition, scored higher on both measures.

One indicator of the validity of this classification comes from the characteristics of the two samples. Whilst it is inevitable that one group will score higher on both tests than the other, it need not be the case that they show the same relative performance on the two measures. In this example, the higher scoring group

were significantly better at non-word reading than exception word reading (Non-words = 22.1, *s.d.* = 5.9; Exception Words = 18.6, *s.d.* = 4.0, $t = -11.14$, $p < 0.001$) but there was no such advantage for the lower scoring (Non-words = 6.7, *s.d.* = 3.9; Exception Words = 18.6, *s.d.* = 4.0, $t = 11.14$, $p < 0.001$). Such a discrepancy in performance suggests that the groups showed a qualitatively different pattern of performance and did not simply differ in overall level of ability. Although this evidence falls short of what would be required to demonstrate that the groups form taxons (Ruscio & Ruscio 2004), it is the sort of difference that might be expected if they did. This difference is also readily interpretable. The high scoring group were significantly stronger at reading non-words than exception words. The natural interpretation of this is that the high scoring group includes those who are well on their way to mastering phonics whereas the lower scoring group has not.

It is notable that although the lower scoring group represents those who are less successful in reading than their peers, the group does not readily map onto dyslexia as generally conceived. First, the proportion of the sample in the lower-scoring group (17.8%) is much higher than the typical prevalence estimates of dyslexia of 5–10% (Rodgers 1983; B. A. Shaywitz, Fletcher, Holahan, & Shaywitz 1992). However, as this was a cross-sectional study, we do not know, whether the reading difficulties seen endured or whether some of the children went on to read well so only some of those identified may have been dyslexic.

In addition, the poor readers form a homogeneous group and there no evidence of the subtypes of dyslexia expected by theory but this, of course, depends upon the sample studied. Quadrant analysis will split the groups into 4 regardless of whether such grouping exist or not and this split will occur at the point of maximal difference between the observed and expected PDFs. In a representative sample of readers, the maximal difference split might well be between skilled and less-skilled. In a sample of dyslexic children, Quadrant Analysis would identify subtypes of dyslexia.

The estimation of the expected PDF was based on the assumption of bivariate normality but, as the marginal distributions of the data were not normally distributed, the joint distribution cannot have been truly bivariate normal, so the assumption was wrong. For this reason, one should not take the attribution of *p*-values in this case too seriously. The assumption of bivariate normality places a major limitation on this approach to selecting optimal cut-off scores because test data is frequently not normally distributed. Only rarely is test data strictly continuous (the possible test scores used here were restricted to integer values from 0 to 30) and there are often floor or ceiling effects (there is a clear ceiling effect on the non-word reading test used here) and both of these are deviations from a strictly normal distribution. As it stands, this means the method described above for choosing cut-off scores is of limited value but it could be revived if a more realistic method for estimating the expected PDF, taking into account deviations of the marginal distributions from the normal, could be devised.

In summary, Quadrant Analysis involves a procrustean imposition of the individual researcher's will on the data and, unsurprisingly, this fails to achieve consensus or validity. For the same reasons, it has no value in identifying taxons. Quadrant Analysis could be improved if there were some explicit, non-arbitrary rationale

for setting the value of threshold used and we propose that the location of the greatest difference between the observed and expected PDFs could be used for this purpose. However, the specific method illustrated here depends upon assumptions of normality that are often unrealistic with cognitive test scores. Despite this limitation, the method does provide an explicit rationale for choosing thresholds and we believe it has heuristic value. Any variant of Quadrant Analysis, however, will share two other serious limitations. First, the method will only be able to identify subtypes that lie within a quadrant; other shaped clusters may be missed. Second, the method is difficult to generalize to more than two or three variables. In short, Quadrant Analysis does not meet the needs of the task in hand and cannot easily be fixed to do so. What is needed is a more objective and rigorous way of identifying subtypes or taxons.

12.5 Alternative Approaches

There are several alternative approaches for identifying latent taxonic structure that offer the prospect of being able to define more objective criteria for identifying distinctive profiles of reading development in children and these include Cluster Analysis (Bonafina, Newcorn, McKay, Koda, & Halperin 2000; King, Giess, & Lombardino 2007; Morris et al. 1998), Finite Mixture Modelling (FMM, McLachlan & Basford 1988; McLachlan & Peel 2000) and Taxometric Analysis (Beauchaine 2003; Meehl 1995). In some senses, each of the methods can be considered as cluster analyses, and Finite Mixture Models often are often classified as such. However, there seems to us to be an important distinction between methods that rely on the pairwise measures of similarity or distance between observations (which we call cluster analysis) and those which generate an explicit parametric model of the data (FMM). The Taxometric approach has its own distinctive history and philosophy of use and, in consequence, deserves separate consideration. One thing that each of these methods has in common, however, and which gives them a considerable advantage over Quadrant Analysis, is that they are all readily generalizable to cases where the subtypes are distinguished by more than two observed variables.

12.5.1 Cluster Analysis

Cluster analysis is a disparate family of methods that identifies groupings in multivariate data sets based on some statistical measure of distance or similarity between observations. There are three main families of cluster analysis: hierarchical clustering, k-means clustering and density based clustering and each of these comes in multiple variations. Hierarchical cluster analysis is particularly fecund and offers a wide choice of similarity (or distance) measures and methods of agglomeration.

For example, SPSS (ver21) offers 37 measures of distance and 7 methods of agglomeration giving 259 possible combinations. Although the choice of distance measure is usually constrained by the type of data being analysed, choosing the appropriate method of agglomeration is more challenging. This abundance would not matter if the methods tended to converge on a consistent result but this is frequently not the case, which makes the decision about which clustering method to use critical.

Both hierarchical and k-means cluster analysis share a common problem. That is, how many clusters should there be? There are multiple methods to help one choose the correct number of clusters including the Akaike Information, Bayes Information, the Calinski-Harabasz and the Davies-Bouldin criteria, and the Silhouette and Gap tests. Unfortunately, these methods often disagree and the question of how many clusters to extract remains unresolved. The third approach to cluster analysis is density-based clustering and unlike the other clustering methods, the number of clusters emerges from the analysis and is not directly pre-determined by the user. However, the user is required to specify other values (critical distance and the minimum cluster size) which, effectively determine the number of clusters so there is no avoiding the issue.

Perhaps surprisingly, cluster analysis has been little used in dyslexia and reading research (Bonafina et al. 2000; King et al. 2007; Morris et al. 1998). Each of these studies primarily addressed the issue of subtypes of children with reading disability and so, unlike Talcott et al. (2002), they did not use a representative sample of all children and only King et al. (2007) compared children with and without reading disability. Each of these studies used k-means clustering (or a variant of it) either alone or with other clustering methods but in other ways (the cognitive measures used and the populations sampled) the studies were very different. Consequently, the character and number of clusters identified were inconsistent. The most sophisticated of these studies King et al. (2007) used cognitive measures derived from theory, state-of-the art criteria for selecting the correct number of clusters and bootstrap sampling to ensure the reliability of the clusters. They found that children without reading disability did not form clusters but that those with reading disability clustered into four groups, consistent with the double deficit hypothesis. As this conclusion was derived from a relatively small sample of 93 children with reading disability, driven by a particular theoretical perspective, the conclusions cannot be considered definitive. Nevertheless, this study, makes a very important contribution to the literature and shows the potential value of cluster analysis in this field of research.

12.5.2 Finite Mixture Modeling

Finite Mixture Modeling (FMM) is a parametric method for identifying clusters in a data set and aims to find the k multivariate PDFs that best account for the observed data. Typically, the PDFs are Gaussian, hence the alternative name for this

approach, Gaussian Mixture Modelling, but other probability distributions can be used if desired. One specific method of note is Latent Class Analysis (LCA) which is a special case of FMM in which the observed variables within each class are uncorrelated.

In conventional FMM, the best fitting combination of PDFs are identified using an expectation maximization algorithm which is equivalent to a maximum likelihood estimate of the parameters of the model. More recently, Bayesian approaches have been introduced that have advantages in robustness and stability of the models and which produce probability distributions of the parameters of the model. In both cases, and like Hierarchical and k-means clustering, the user needs to specify the number of clusters in advance. Unlike hierarchical and k-means clustering, which allocate each observation to a single cluster (“hard” clustering), each individual observation will have a given probability of belong to each of the k clusters defined by the cluster’s PDF (“soft” clustering). Individual observations are allocated to the cluster that has the maximum probability for that case.

As far as we are aware, FMM has not been used to identify subtypes of reading disability or to separate typical from atypical readers. As FMM has some advantages over other forms of cluster analysis, this is an omission that should be corrected.

12.5.3 *Taxometric Analysis*

The final approach to identifying subtypes in dyslexia that we will consider here is taxometric analysis. Taxometrics is a term used to describe a family of methods developed by Paul Meehl and colleagues for determining whether a multivariate data set consists of a latent taxonic structure or not (Beauchaine 2007; Meehl 1995; Ruscio, Haslam, & Ruscio 2006). Taxometrics consists of several distinct methods known as “*coherent cut kinetics*”, that each seek to identify abrupt discontinuities in what appear to be continuous parameters of the measures that distinguish putative members of the taxon in question from non-taxon members. These key measures are referred to as “indicator” variables and the existence of a discontinuity between taxon and non-taxon group members on the indicator variables is taken as evidence for taxonic structure. There are five distinct methods known as MAXSLOPE, MAMBAC, L-Mode, MAXCOV AND MAXEIG that each look for a discontinuity in a different parameter (local regression slope, local mean, latent factor, covariance and eigenvalues, respectively). Taxometrics places a strong emphasis on converging evidence from across these different methods.

Taxometrics has many advantages over other approaches in that it is objective, quantifiable and uses convergent evidence to establish taxonic structure. However, large sample sizes (average size ~600) and large effect sizes (Cohen’s $d > 2$) are needed. Taxometric methods are insensitive in cases where the taxon makes up only a small proportion of the total sample (<15%) and are susceptible to sampling bias and distributional skew in the indicator variables. In addition, the methods do not work well when there are substantial within-group correlations ($r > 0.4$) between

the indicator variables. Despite these limitations, taxometric methods have made significant contributions to the classification of many adult mental health disorders and have been notably successful in delineating the relationship between personality and adult psychopathology (see Chap. 10, Ruscio et al. 2006).

Only one study has used this approach with dyslexia (O'Brien, Wolf, & Lovett 2012). Using a large sample of 671 children with severe reading disorders aged 6–8 years old assessed on a range MAMBAX, MAXEIG and L-mode (Ruscio et al. 2006). They concluded that there was evidence of two taxa of dyslexia, those with and without phonological deficits. This approach appears most promising and merits replication and extension.

12.6 Discussion

There are two important and related questions that we address here. First, do children with dyslexia show fundamental and enduring non-arbitrary and objective differences from children with typical reading profiles? Second, is dyslexia a single condition or does it consist of multiple subtypes, each differing in some fundamental and enduring, non-arbitrary way from the others. In the language of taxometrics, these questions are whether the variation in developmental reading profiles can best be considered as dimensional or taxonic. Our objective was not to answer these questions, but to consider ways in which these question might be answered.

We showed that the common practice of performance thresholds on one or more measures of cognitive ability to delineate dyslexics from the typical readers, a method we refer to as Quadrant Analysis, can never hope to answer this question. Quadrant Analysis, as the research literature attests, is essentially arbitrary and will identify subtypes of readers regardless of whether they exist in any fundamental or objective sense. As such, it simply will not serve to answer the question as to whether we are dealing with dimensions or taxons.

Despite this, the temptations of simplicity of concept and ease of use make it likely that Quadrant Analysis will continue to be used. Acknowledging this, we proposed a modification to Quadrant Analysis that makes the choice of threshold less arbitrary and which might prove useful on occasion. We do not claim, however, that this will overcome all the problems of Quadrant Analysis, let alone resolve the “*category or dimension?*” issue.

Instead, we believe that alternative approaches, including cluster analysis, FMM and taxometrics will prove more useful. Unlike Quadrant Analysis, each of these approaches can be readily adapted to more than two observed variables, and are probably more powerful when used this way. However, these methods are not panaceas and each has limitations and presents the potential user with their own unique challenges. With all such methods, the output depends upon the input. The choice of observed measures, indicator measures in the parlance of taxometrics, is particularly critical. Measures that show some ability to discriminate between the putative subtypes should be preferred and the inclusion of too many

irrelevant measures can obscure real differences. Contrariwise, too many highly correlated variables, whether they discriminate individually or not, simply increases redundancy in the model and can impair the ability of the statistical algorithms to reach a stable solution. The sampling strategy is also critical. Choosing a representative sample of learner readers might be useful for answering the “*Is dyslexia a taxon?*” question, but would be of little use in identifying subtypes of dyslexia as too few individuals of each subtype would be present in the sample. To address this question, a representative sample of problem readers would be more useful. Whichever question is addressed, large samples sizes of several hundred cases will be required to give reliable results.

To date, attempts to use these methods with reading profiles have been few and far between with a mere handful of studies available in the published literature (Bonafina et al. 2000; King et al. 2007; Morris et al. 1998; O’Brien et al. 2012). This is a shame but one that could be corrected quite easily. There are multiple databases of problem readers and representative samples of learner readers that could be used. There are also multiple theories about the nature of dyslexia to guide the choice of indicator variables. This being so, the application of cluster analysis, FMM and taxometrics to these important questions should be straightforward.

No matter what methods we use, however, we should not expect a rapid resolution to the taxonomy of dyslexia question. The much-delayed publication of DSM-V (American Psychiatric Association 2013), where the “*category or dimension?*” issue was the focus of debate around several mental disorders, does not provide grounds for optimism. In none of these cases can the issue be considered closed, including those where the evidence base from cluster analysis and taxometrics is much better established than it is with dyslexia (e.g., schizophrenia). As for dyslexia itself, the wide range of disparate views, strongly held opinions and absence of evidence one way or the other makes consensus seem a distant destination. Nevertheless, the accumulation of evidence derived from statistical tools specifically designed to address the “*category or dimension?*” question seems to us to be a good place to start.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Washington, DC: American Psychiatric Publishing.
- Badian, N. A. (2005). Does a visual-orthographic deficit contribute to reading disability? *Annals of Dyslexia*, 55(1), 28–52. <https://doi.org/10.1007/s11881-005-0003-x>
- Beauchaine, T. P. (2003). Taxometrics and developmental psychopathology. *Development and Psychopathology*, 15(3), 501–527. <https://doi.org/10.1017/S0954579403000270>
- Beauchaine, T. P. (2007). A brief taxometrics primer. *Journal of Clinical Child and Adolescent Psychology*, 36(4), 654–676. <https://doi.org/10.1080/15374410701662840>
- Boder, E. (1973). Developmental dyslexia: A diagnostic approach based on three atypical reading-spelling patterns. *Developmental Medicine & Child Neurology*, 15(5), 663–687. <https://doi.org/10.1111/j.1469-8749.1973.tb05180.x>

- Bonafina, M. A., Newcorn, J. H., McKay, K. E., Koda, V. H., & Halperin, J. M. (2000). ADHD and reading disabilities: A cluster analytic approach for distinguishing subgroups. *Journal of Learning Disabilities, 33*(3), 297–307. <https://doi.org/10.1177/002221940003300307>
- Bosse, M.-L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition, 104*(2), 198–230. <https://doi.org/10.1016/j.cognition.2006.05.009>
- Botev, Z. I. (2015). *Kernel density estimation*. <https://uk.mathworks.com/matlabcentral/fileexchange/17204-kernel-density-estimation>
- Botev, Z. I., Grotowski, J. F., & Kroese, D. P. (2010). Kernel density estimation via diffusion. *The Annals of Statistics, 38*(5), 2916–2957. <https://doi.org/10.1214/10-Aos799>
- Bowers, P. G., & Wolf, M. (1993). Theoretical links among naming speed, precise timing mechanisms and orthographic skill in dyslexia. *Reading and Writing, 5*(1), 69–85. <https://doi.org/10.1007/Bf01026919>
- Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47*(2), 149–180. [https://doi.org/10.1016/0010-0277\(93\)90003-E](https://doi.org/10.1016/0010-0277(93)90003-E)
- Castles, A., Datta, H., Gayan, J., & Olson, R. K. (1999). Varieties of developmental reading disorder: Genetic and environmental influences. *Journal of Experimental Child Psychology, 72*(2), 73–94. <https://doi.org/10.1006/jecp.1998.2482>
- Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). London: Academic Press.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256. <https://doi.org/10.1037/0033-295X.108.1.204>
- Elliot, C., Murray, D., & Pearson, L. (1983). *British ability scales*. Windsor, UK: National Foundation for Educational Research.
- Elliott, J., & Grigorenko, E. L. (2014). *The dyslexia debate* (Vol. 14). New York, NY: Cambridge University Press.
- Genard, N., Mousty, P., Content, A., Alegria, J., Leybaert, J., & Morais, J. (1998). Methods to establish subtypes of developmental dyslexia. In P. Reitsma & L. T. Verhoeven (Eds.), *Problems and interventions in literacy development* (Vol. 15, pp. 163–176). Dordrecht: Kluwer Academic Publishers.
- Jimenez, J. E., Rodriguez, C., & Ramirez, G. (2009). Spanish developmental dyslexia: Prevalence, cognitive profile, and home literacy experiences. *Journal of Experimental Child Psychology, 103*(2), 167–185. <https://doi.org/10.1016/j.jecp.2009.02.004>
- King, W. M., Giess, S. A., & Lombardino, L. J. (2007). Subtyping of children with developmental dyslexia via bootstrap aggregated clustering and the gap statistic: Comparison with the double-deficit hypothesis. *International Journal of Language & Communication Disorders, 42*(1), 77–95. <https://doi.org/10.1080/13682820600806680>
- Manis, F. R., Seidenberg, M. S., Doi, L. M., McBride-Chang, C., & Petersen, A. (1996). On the bases of two subtypes of development dyslexia. *Cognition, 58*(2), 157–195. [https://doi.org/10.1016/0010-0277\(95\)00679-6](https://doi.org/10.1016/0010-0277(95)00679-6)
- McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering* (Vol. 84). New York: Dekker.
- McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. Hoboken, NJ: John Wiley & Sons, Inc. <https://doi.org/10.1002/0471721182>
- Meehl, P. E. (1995). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50*(4), 266–275. <https://doi.org/10.1037//0003-066x.50.4.266>
- Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., ... Shaywitz, B. A. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology, 90*(3), 347–373. <https://doi.org/10.1037/0022-0663.90.3.347>
- O'Brien, B. A., Wolf, M., & Lovett, M. W. (2012). A taxometric investigation of developmental dyslexia subtypes. *Dyslexia, 18*(1), 16–39. <https://doi.org/10.1002/dys.1431>

- Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, 101(2), 385–413. <https://doi.org/10.1016/j.cognition.2006.04.008>
- Peterson, R. L., Pennington, B. F., Olson, R. K., & Wadsworth, S. (2014). Longitudinal stability of phonological and surface subtypes of developmental dyslexia. *Scientific Studies of Reading*, 18(5), 347–362. <https://doi.org/10.1080/10888438.2014.904870>
- Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly*, 27(1), 28. <https://doi.org/10.2307/747832>
- Rodgers, B. (1983). The identification and prevalence of specific reading retardation. *British Journal of Educational Psychology*, 53(3), 369–373. <https://doi.org/10.1111/j.2044-8279.1983.tb02570.x>
- Ruscio, J., Haslam, N., & Ruscio, A. M. (2006). *Introduction to the taxometric method: A practical guide*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ruscio, J., & Ruscio, A. M. (2004). Clarifying boundary issues in psychopathology: The role of taxometrics in a comprehensive program of structural research. *Journal of Abnormal Psychology*, 113(1), 24–38. <https://doi.org/10.1037/0021-843X.113.1.24>
- Shaywitz, B. A., Fletcher, J. M., Holahan, J. M., & Shaywitz, S. E. (1992). Discrepancy compared to low achievement definitions of reading disability: Results from the Connecticut Longitudinal Study. *Journal of Learning Disabilities*, 25(10), 639–648. <https://doi.org/10.1177/002221949202501003>
- Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *The New England Journal of Medicine*, 326(3), 145–150. <https://doi.org/10.1056/NEJM199201163260301>
- Sprenger-Charolles, L., Colé, P., Lacert, P., & Serniclaes, W. (2000). On subtypes of developmental dyslexia: Evidence from processing time and accuracy scores. *Canadian Journal of Experimental Psychology*, 54(2), 87–104. <https://doi.org/10.1037/h0087332>
- Sprenger-Charolles, L., Siegel, L. S., Jiménez, J. E., & Ziegler, J. C. (2011). Prevalence and reliability of phonological, surface, and mixed profiles in dyslexia: A review of studies conducted in languages varying in orthographic depth. *Scientific Studies of Reading*, 15(6), 498–521. <https://doi.org/10.1080/10888438.2010.524463>
- Talcott, J. B., Witton, C., Hebb, G. S., Stoodley, C. J., Westwood, E. A., France, S. J., ... Stein, J. F. (2002). On the relationship between dynamic visual and auditory processing and literacy skills: Results from a large primary-school study. *Dyslexia*, 8(4), 204–225. <https://doi.org/10.1002/dys.224>
- Talcott, J. B., Witton, C., & Stein, J. F. (2013). Probing the neurocognitive trajectories of children's reading skills. *Neuropsychologia*, 51(3), 472–481. <https://doi.org/10.1016/j.neuropsychologia.2012.11.016>
- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 101(2), 192–212. <https://doi.org/10.1037/0033-2909.101.2.192>
- Ziegler, J. C. (2008). Better to lose the anchor than the whole ship. *Trends in Cognitive Sciences*, 12(7), 244–5. <https://doi.org/10.1016/j.tics.2008.04.001>