Screening for Dyslexia, Dyspraxia and Meares-Irlen Syndrome in Higher Education

S. A. Nichols¹, J. S. McLeod¹, R. L. Holder² and H. S. T. McLeod²

¹Dyslexia and Disability Service, University of Worcester, Worcester, UK
²School of Health and Population Sciences, University of Birmingham, Birmingham, UK

This study reports a comparison of screening tests for dyslexia, dyspraxia and Meares-Irlen (M-I) syndrome in a Higher Education setting, the University of Worcester. Using a sample of 74 volunteer students, we compared the current tutor-delivered battery of 15 subtests with a computerized test, the Lucid Adult Dyslexia Screening test (LADS), and both of these with data on assessment outcomes. The sensitivity of this tutor battery was higher than LADS in predicting dyslexia, dyspraxia or M-I syndrome (91% compared with 66%) and its specificity was lower (79% compared with 90%). Stepwise logistic regression on these tests was used to identify a better performing subset of tests, when combined with a change in practice for M-I syndrome screening. This syndrome itself proved to be a powerful discriminator for dyslexia and/or dyspraxia, and we therefore recommend it as the first stage in a two-stage screening process. The specificity and sensitivity of the new battery, the second part of which comprises LADS plus four of the original tutor delivered subtests, provided the best overall performance: 94% sensitivity and 92% specificity. We anticipate that the new two-part screening process would not take longer to complete. Copyright © 2008 John Wiley & Sons, Ltd.

Keywords: dyslexia; dyspraxia; Meares-Irlen syndrome; higher education; screening; visual stress
INTRODUCTION

Universities have a responsibility to ‘prevent disabled people from being treated less favourably than others’ (DDA, 1995; S29(2)), and 50% of those disclosing a disability in Higher Education (HE) are dyslexic (HESA, 2007). Given the recently published requirement of the Disability Equality Duty (Disability Rights Commission, 2007) that all institutions should be pro-active in their efforts to ensure equality, it is important to determine mechanisms that can sensitively and economically screen students for a range of disabilities including dyslexia, in order to identify those who require full assessment, the gateway to support. A crucial decision for every HE institution in achieving this end is whether to opt for screening by computer or by tutor.

This research analyses the effectiveness of the current tutor screening practice at the University of Worcester (UW) by comparison with a computerised screening tool. The performance of the two screening instruments for detecting learning disabilities was examined in a prospective sample obtained through the university Disability and Dyslexia Service (DDS). We collected data on the original tutor-delivered battery of 15 subtests, which was being used at UW to screen for dyslexia, dyspraxia and Meares-Irlen syndrome (M-I syndrome). We screened the same participants using a computerized screening tool for dyslexia, the Lucid Adult Dyslexia Screening (LADS) test (Singleton, 2004b). In addition, we collected data from these participants’ assessments, and used these data to explore the possibility of a better battery for screening for these three targeted disabilities.

There has been some concern about adult screening efficacy among Specific Learning Disabilities (SpLD) professionals: the 1999 working party declared that ‘nationally adopted higher education-related identificatory screening test batteries, given before sending students for a full assessment, could help improve the consistency of both the individual and institution’s decision-making in the identification of dyslexia and the award of DSA [Disabled Students' Allowance]’ (Singleton, 1999, p. 51). This was still a driver for the 2005 Working Group (DfES, 2005), against a background of a ‘paucity of appropriate screening procedures for use with the suspected adult dyslexic’ (Kirk, McLoughlin, & Reid, 2001, p. 28).

Many dyslexic students are not diagnosed until they enter HE (40% according to Singleton (2004a)), and our experience at UW is that many of these remain unidentified and unsupported until their final year. Given the scarcity of screening tools for dyspraxia and M-I syndrome, this probably also holds true for those with these conditions. Presumably, some students with SpLD graduate, or drop out before graduating, without being identified.

Our aims were to quantify the difference in performance between the Tutor Screening Battery (TSB) we were currently employing and that of LADS, as indicators of dyslexia, dyspraxia and M-I syndrome, and to determine whether a more powerful screening instrument could be devised.

DEFINITIONS

This paper uses the definition of dyslexia developed by the 1999 working party: ‘dyslexia is evident when accurate and fluent word reading and/or spelling
develops very incompletely or with great difficulty. This focuses on literacy learning at the ‘word level’ and implies that the problem is severe and persistent despite appropriate learning opportunities’ (Singleton, 1999, p. 18).

Dyspraxia is here regarded as an impairment or immaturity of the organization of movement. This may be associated with problems of language, perception and thought (Dyspraxia Foundation, 2007). The identification of dyspraxia is ultimately a medical diagnosis. In this paper, the term means ‘SpLD consistent with a diagnosis of dyspraxia’.

In identifying M-I syndrome we follow the definition by Kriss and Evans (2005, p. 1) of a syndrome characterized by ‘symptoms of visual stress and visual perceptual distortions that are alleviated by using individually prescribed coloured filters’. M-I syndrome is not currently defined as a specific learning difficulty, but is known to occur more frequently among dyslexics than in the general population (Kriss & Evans, 2005; Singleton & Trotter, 2005), and certainly hampers full or easy access to any HE curriculum. Readers with M-I syndrome read faster, or in greater comfort, when using a coloured acetate overlay or coloured lenses than when reading black text on a white background.

ORGANIZATIONAL CONTEXT

The research was conducted within an operational setting: all five members of the research team work for the DDS at UW. At UW, approximately 120 students per annum currently refer themselves to the Disability and Dyslexia Service (UW, 2008), asking whether they should be assessed, often on the advice of a tutor or because of long-standing anxieties. Dyslexia has the highest proportion of declared disability at UW, at 53% of all disabled students (UW, 2008). The screening system for predicting the three disabilities at UW relies on self-referral followed by screening by trained tutors; no pro-active screening takes place. M-I syndrome tests are not currently offered at UW until the assessment stage.

Our purpose in screening students is to identify as many students with the targeted disabilities as possible and to exclude those without (Harrison & Nichols, 2005; Payne, 1998). This means that screening needs to be sensitive to a range of SpLDs and not simply to the most common, dyslexia. This forms the rationale for the screening of students at UW, and this is why we consider the performance of published screening instruments for conditions for which they were not designed.

METHODOLOGY

The Tutor Screening Battery

This section explains our choice of subtests in the battery. As screeners, we are looking for signs of the same difficulties used by assessors to identify SpLD for the purpose of providing appropriate support and/or application for the DSA:

- working memory;
- phonological processing;
visual processing;
- sequencing and orientation;
- hand–eye coordination;
- spelling;
- reading;
- writing

plus any genetic factors.

The TSB is a combination of parts of the Bangor Dyslexia Test (BDT) (Miles, 1997) hereafter Bangor Dyslexia Test Abridged (BDTA), and parts of the Dyslexia Adult Screening Test (DAST) (Fawcett & Nicolson, 1998), hereafter Dyslexia Adult Screening Test Abridged (DASTA). Students were referred to assessment on reaching a predetermined cutoff on either group of subtests.

The BDT (Miles, 1997)

We are not aware of this test being normed on an adult population. It was originally intended as a brief assessment to identify dyslexia in children (Van Daal & Miles, 2001), but has latterly been used extensively as a screening test. Devised in the early 1980s, before much attention had been focused on phonological processing deficit, its subtests chiefly emphasize difficulties of working memory, sequencing and orientation, providing for example an objective measure of left/right confusion. It also asks two questions; one re familial incidence and one re b/d muddles. After experimentation during our pilot study, we decided to use the digit span test from DAST but not from BDT.

Thus, we administered all of the BDT subtests except for digit span (forwards and reversed): left/right confusions, polysyllable repetition, subtraction, times tables, months forwards, months reversed, b/d confusions and familial incidence. BDT uses a three-point scoring system, 0, 0.5 or 1, where 0 equates to success at the test, 0.5 equates to satisfying some of the criteria and 1 equates to failing to satisfy the criteria. Our cutoffs for full assessment were 2.5 positive indicators on our selected subtest battery of eight, BDTA.

The DAST (Fawcett & Nicolson, 1998)

DAST draws from work on phonological deficit (Frith, 1999; Snowling, 1995), and on the cerebellar hypothesis (see, for example, Fawcett, 2001; Nicolson & Fawcett, 1999). It has proved very successful in HE: ‘the data it provides is more than one would expect from a screening test’ (Reid & Kirk, 2001, p.33). Unlike the BDT, it was devised as a screening test from the outset, and provides an ‘At Risk’ measure for dyslexia.

Harrison and Nichols (2005) found that overall it had a sensitivity of 74% and a specificity of 84%. They found that the subtests that were most reliable at discriminating between dyslexics and non-dyslexics were nonsense passage reading, two minute spelling, phonemic segmentation, one minute reading and one minute writing. They also maintained that the postural stability, nonverbal reasoning and semantic fluency subtests were not effective discriminators. Gunn (2000) also found that semantic fluency and nonverbal reasoning did not distinguish between dyslexic adults and those with general learning difficulties. In a
study using similar measures to DAST, Hatcher, Snowling, and Griffiths (2002) found that the best predictors were nonword reading, spelling, digit span and writing speed.

Landerl, Frith, and Wimmer (1996) found phonemic difficulties to be a central problem in dyslexia. We therefore decided that the balance of evidence was in favour of including the phonemic segmentation subtest. It was our own experience that this is a particularly good discriminator for well-compensated dyslexic students. We also decided to include rapid naming, as this seemed to us to be a valuable measure of the students’ ability to access and articulate their internal lexicon, and might be an indicator of dyspraxia as well as of difficulties in phonological processing.

We had some reservations about the use of the postural stability subtest on students in a confined space. Moreover, a recent meta-analysis found that impaired balance did not have a specific enough association with dyslexia to be used in identification (Rochelle & Talcott, 2006). We therefore excluded this subtest from our battery.

The more of the DAST subtests employed, the higher their combined predictive value, according to Fawcett and Nicolson (1998). We decided to use all the subtests whose efficacy was demonstrated either by Harrison and Nichols, or by Hatcher et al., plus rapid naming. This group of seven subtests is referred to here as the DASTA.

DAST uses a four-point scoring system, 0, 1, 2 or 3. Zero equates to the student performing within the normal range on that task, and 1, 2 and 3 represent points successively below that normal range. These scores are achieved by comparing the raw score with age-related score keys for population norms. The seven subtests taken together give a partial quantitative ‘At Risk Quotient’ (ARQ) (Fawcett & Nicolson, 1998), by dividing the students’ total score by seven (the number of subtests). Our cutoffs for full assessment were six or more on DASTA. This meant that our ARQ, the cutoff warranting referral for assessment, was 0.857 (6/7), rather than the original DAST recommended ARQ of 1. This necessarily gave the DASTA group of subtests a higher sensitivity than would be expected when using the full DAST.

Thus, altogether the TSB comprised eight subtests of BDT and seven of DAST.

The LADS Screening Battery
LADS employs an algorithm that weights performance on its nonverbal reasoning test against performance on its subtests of word recognition, word construction and memory. In other words, the student who scores highly on reasoning will have to do worse on the other subtests to be identified as at high risk of dyslexia than the student who does less well on reasoning.

LADS gives a risk result for each element and a brief summary identifying the student as at high, moderate, borderline or low risk of dyslexia. Like BDT and DAST, it is not intended or designed to indicate the presence of any other SpLD, nor of M-I syndrome. We referred for assessment all those identified by LADS as at borderline risk of dyslexia or above, as recommended in the LADS manual (Singleton, Horne, & Thomas, 2004) and verified by our pilot study.
Pilot Study

We conducted a small pilot study in order to

1. analyse and improve inter-rater agreement (Goodwin, 2001);
2. assess cutoff points used, using likelihood ratios, sensitivity and specificity, to
determine whether they were appropriate.

The pilot study indicated that the cutoffs being used for the two parts of the
TSB were appropriate and also that the order of delivery of screening test had no
discernible effect on the result of the test or the students’ preference for one or the
other; TSB or LADS.

Main Study

Recruitment and Participants

We presented details of our research to first, second and third year cohorts
of students on BA courses in Primary Initial Teacher Education at UW, and
invited them to participate, offering a small financial inducement. Attempts to
recruit from the nursing department were abandoned early as they proved
difficult to access and appeared to be volunteering on a self-referral basis: all of
the first five proved positive for dyslexia at assessment. Data for two of these
participants were complete and are included in the analysis. We accepted
volunteers even if they had had previous assessments, as we wanted to
understand the characteristics of the population. One hundred students were
initially recruited and 74 were ultimately included in the study. The average age
of these was 23.4 years. Ninety-two percent of the 74 were female and 8% male
(see Figure 1).

Screening Procedure

Individual screening times were arranged at the convenience of each student for
the tutor screening, with group screening times being offered for LADS. Thus, the
order of delivery of screening was randomised. Screenings were conducted blind:
no screener or participant knew the result of the first screening at the second.
When any participant withdrew, all of their data were removed from the dataset.

Screenings were administered between May 2007 and January 2008. The
shortest time between TSB and LADS screening was a few hours; the longest was
six and a half months. The shortest time from second screening to assessment
was 3 days, and the longest was 7 months.

Assessment Procedure

We needed to replicate the criteria currently used in HE to choose assessors. So
we chose

1. only assessors qualified to assess adults for the DSA (all assessments
   conducted and reports written in accordance with the DfES guidelines
   for the assessment of SpLDs in HE (DfES, 2005));
2. assessors who only used tests from the approved list set out by Patoss (2007).
Seven qualified assessors were employed. Six held current practising certificates in the assessment of SpLDs; the seventh was a chartered educational psychologist. Three are members of the research team.

In accordance with current policy at UW, assessors were given the results of both screenings in advance. This gave the assessor a useful starting point, and also enabled an informed judgement about whether either of the screenings was more useful to them (i.e. saved them time or provided valuable information, either of which might have cost implications). If so, these differences will be taken into account in any future cost comparison.

We requested all assessors to take particular note of dyspraxia and M-I syndrome. If assessors suspected dyspraxic tendencies, we asked them to test the student using the Morrisby Test of Manual Dexterity (1991) and/or the Beery-Buktenica Developmental Test of Visual–Motor Integration (Beery & Beery, 2006), and the Dyspraxia Foundation’s checklist (2007) for symptoms of dyspraxia in adults (see the Guidance from the SpLD Working Group (DfES, 2005)).

We asked assessors to test for M-I syndrome using Cerium coloured overlays (Cerium Visual Technologies (2007) and the Wilkins Rate of Reading Test (Wilkins, 1996). Assessors did not attempt to identify this on the basis of sustained use of an overlay, but simply on the basis of immediate improvement in reading speed of over 5% on the Wilkins Rate of Reading Test (Wilkins, Jeanes, Pumfrey, & Laskier, 1996) when using a preferred overlay. Not all assessors in fact acceded to this request; where the test was not administered in full we withdrew the participant’s data from the dataset (14 cases). Those students who were fully tested and whose reading speed improved by 5% or more were referred to optometrists specializing in this field. Students reporting increased ease of reading, without showing a 5% increase in reading speed, were advised to use their preferred overlay and to experiment with changing the background colour of their computer screen. Findings from the optometrists consulted were not available for this paper.

**Statistical Analysis**

Logistic regression was undertaken to identify the association between each individual subtest employed in the TSB and the target disabilities at assessment. Logistic regression was also used to compare the overall performance of the TSB and LADS. The results of each regression analysis were reported using adjusted odds ratios (with 95% confidence intervals) and p values. The Nagelkerke R squared was also reported, where appropriate, as a measure of each model’s performance.

Analysis was also undertaken to identify an enhanced screening battery. This proved to be a two-step process. The first step would be to screen students for M-I syndrome. The second step would be to apply a new battery of screening subtests on students who screened negative for M-I syndrome. Stepwise logistic regression was also undertaken to identify the strongest performing set of subtests associated with dyslexia and/or dyspraxia, using data from students who had been found not to have M-I syndrome. The co-efficients from this analysis were used to identify alternative combinations of sensitivity and specificity for the enhanced screening battery as a whole. This entailed using the following formula to calculate \( Y \), the probability of a student found free of M-I syndrome...
syndrome having dyslexia and/or dyspraxia.

\[
Y = \frac{1}{1 + \exp(-X)}
\]

where \(X = a_1 x_1 + a_2 x_2 + \cdots (x_1, x_2 \ldots\) represent the scores on individual test items and \(a_1, a_2 \ldots\) are numbers chosen to give the best predictor of \(Y\) from \(x_1, x_2 \ldots\).

MINITAB was then used to calculate the probability ‘cutoff’ associated with the optimal combination of sensitivity and specificity. Sensitivity and specificity combinations were then calculated for the enhanced screening test as a whole.

The odds ratio shows the ratio of the odds of the presence of the disability with a positive screening test compared with the odds with a negative screening test. The adjusted odds ratio takes into account the information from the other screening tests included in the analysis. Stepwise logistic regression identifies the most efficient combination of subtests; individual subtests may be excluded from this combination because they are correlated with other subtests (i.e. duplicating work done by another subtest). Subtests with comparatively less significant \(p\) values may be included in the combination because they are associated with (i.e. ‘pick up’) students who would otherwise be missed.

RESULTS

Participants

One hundred students were recruited. One student withdrew before completing both screening tests, and seven before being assessed, despite having received positive indicators in at least one screening. Four students were excluded from the analysis because they had not yet been assessed. A further 14 students were excluded from the analysis because their assessment did not include the test for M-I syndrome (see Figure 1). Thus, of the 100 students initially recruited, 74 were finally included in the study. Comparison of the 14 students excluded from the analysis indicated no differences in terms of age or the route they took into HE.

Screening Results

Sixty-six percent (49/74) of the students received the same screening outcome from both the TSB (scoring 2.5 or more on BDTA or 6 or more on DASTA) and LADS (scoring ‘high’, ‘moderate’ or ‘borderline’ risk). Twenty-eight per cent (21/74) of participants were screened positive by both screening methods, and 38% (28/74) were double negatives. Eight percent (6/74) were positive on the LADS screening and negative on the TSB, while 26% (19/74) were positive on TSB and negative on LADS. These results, and their equivalents for the excluded participants, are shown in Table 1.

Assessment Results

As outlined above, 74 students were included in the analysis. Sixty-two per cent (46/74) of these were screened positive, and were referred for full assessment. The remaining 28 had received negative results from both screenings, but were
nonetheless invited to attend a full assessment. Ten volunteered, all of whom were given full assessments and all of whom were found to be negative for dyslexia, dyspraxia and M-I syndrome. On this basis, we decided to include the other 18 in our analysis as ‘true’ negatives (Figure 1).

Of the 74 students included in the analysis, 47% (35/74) were found to have at least one of the disabilities under investigation, and 86% (30/35) of those

---

**Table 1. Characteristics of sample and exclusions due to missing data on M-I syndrome**

<table>
<thead>
<tr>
<th></th>
<th>Students included in analysis (n = 74)</th>
<th>Students excluded because of missing data on M-I syndrome (n = 14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% (Number) dyslexic</td>
<td>41 (30)</td>
<td>64 (9)</td>
</tr>
<tr>
<td>% (Number) dyspraxic</td>
<td>20 (15)</td>
<td>21 (3)</td>
</tr>
<tr>
<td>% (Number) positive Tutor Screening Battery(^a)</td>
<td>54 (40)</td>
<td>86 (12)</td>
</tr>
<tr>
<td>% (Number) positive LADS screening(^a)</td>
<td>36 (27)</td>
<td>57 (8)</td>
</tr>
</tbody>
</table>

\(^a\)For dyslexia, dyspraxia or M-I.
students were dyslexic. All 30 were declared to be eligible for the DSA. There was considerable overlap (co-morbidity) between conditions (Figure 2): the majority (57%, 20/35) of students with any one of these disabilities proved to have two or all three. Twenty percent (15/74) were assessed as dyspraxic, using the criteria outlined above. Four percent (3/74) showed dyspraxia alone. M-I syndrome was identified in 24% (18/74). Only one student (1%, 1/74) was found to have M-I syndrome alone, without either dyslexia or dyspraxia.

**Tutor Screening Subtests**

Logistic regression on the BDTA and DASTA groups of subtests was undertaken where students scoring a total of 2.5 or more on BDTA and/or 6 or more on DASTA were assigned a value of 1 and all those not achieving this threshold were assigned a value of 0. The $R^2$ squared for this model was 0.638, and the performance of each group of subtests is summarized in Table 2.

A logistic regression analysis between each individual subtest employed in the TSB and the identification of any of the three disabilities at assessment was

![Figure 2. Co-morbidity of dyslexia, dyspraxia and Meares-Irlen syndrome from the 74 students included in the analysis.](image)

**Table 2.** Logistic regression results: odds ratios and $p$ values of association for the BDTA and DASTA groups of subtests with dyslexia and/or dyspraxia and/or MI syndrome

<table>
<thead>
<tr>
<th>Group of subtests</th>
<th>Adjusted odds ratio (OR)</th>
<th>$p$ Value</th>
<th>95% Confidence interval for adjusted odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower band</td>
</tr>
<tr>
<td>BDTA</td>
<td>24.354</td>
<td>$&lt;0.001^{**}$</td>
<td>5.749</td>
</tr>
<tr>
<td>DASTA</td>
<td>10.095</td>
<td>0.002**</td>
<td>2.345</td>
</tr>
</tbody>
</table>

$^{**}p<0.01$, $^*p<0.05$.  

DOI: 10.1002/dys
undertaken in order to ascertain the individual association between each subtest and any of the three disabilities. The results are summarized in Table 3. The $R^2$ squared for this model was 0.791.

**TSB Versus LADS**

Logistic regressions on the TSB and LADS were undertaken. Students scoring 2.5 or more on BDTA and/or 6 or more on DASTA were assigned a value of 1 for the TSB, and all those not achieving this threshold were assigned a value of 0. Students scoring ‘high’, ‘moderate’ or ‘borderline’ risk for LADS were assigned a value of 1, and those scoring ‘low’ risk were assigned a value of 0. The performance of each of the screening tests is summarized in Table 4.

### Table 3. Adjusted odds ratios and $p$ values of association between subtests and with dyslexia and/or dyspraxia and/or MI syndrome

<table>
<thead>
<tr>
<th>Group of subtests</th>
<th>Subtest</th>
<th>Adjusted odds ratio (OR)</th>
<th>$p$ Value</th>
<th>95% Confidence interval for adjusted odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower band</td>
</tr>
<tr>
<td>BDTA</td>
<td>Left/right confusion</td>
<td>26.513</td>
<td>0.019*</td>
<td>1.713</td>
</tr>
<tr>
<td>BDTA</td>
<td>Polysyllables</td>
<td>13.735</td>
<td>0.076</td>
<td>0.764</td>
</tr>
<tr>
<td>BDTA</td>
<td>Subtraction</td>
<td>11.656</td>
<td>0.200</td>
<td>0.273</td>
</tr>
<tr>
<td>BDTA</td>
<td>Tables</td>
<td>4.874</td>
<td>0.182</td>
<td>0.475</td>
</tr>
<tr>
<td>BDTA</td>
<td>Months reversed</td>
<td>1506.912</td>
<td>0.042*</td>
<td>1.311</td>
</tr>
<tr>
<td>BDTA</td>
<td>b/d Confusion</td>
<td>2.893</td>
<td>0.573</td>
<td>0.072</td>
</tr>
<tr>
<td>BDTA</td>
<td>Familial incidence</td>
<td>2.014</td>
<td>0.610</td>
<td>0.136</td>
</tr>
<tr>
<td>DASTA</td>
<td>Rapid naming</td>
<td>2.870</td>
<td>0.188</td>
<td>0.597</td>
</tr>
<tr>
<td>DASTA</td>
<td>One minute reading</td>
<td>11.308</td>
<td>0.011*</td>
<td>1.759</td>
</tr>
<tr>
<td>DASTA</td>
<td>Phonemic seg.</td>
<td>2.915</td>
<td>0.022*</td>
<td>1.164</td>
</tr>
<tr>
<td>DASTA</td>
<td>Two minute spelling</td>
<td>9.422</td>
<td>0.032*</td>
<td>1.200</td>
</tr>
<tr>
<td>DASTA</td>
<td>Digit span</td>
<td>3.572</td>
<td>0.082</td>
<td>0.853</td>
</tr>
<tr>
<td>DASTA</td>
<td>Nonsense passage reading</td>
<td>0.175</td>
<td>0.066</td>
<td>0.028</td>
</tr>
<tr>
<td>DASTA</td>
<td>One minute writing</td>
<td>0.814</td>
<td>0.759</td>
<td>0.218</td>
</tr>
</tbody>
</table>

**$p<0.01$, *$p<0.05$. Months forward is excluded from the table due to insufficient data (only one student screened positive with this test).**

### Table 4. Logistic regression results: odds ratios and $p$ values of association for the Tutor Screening Battery and LADS with dyslexia and/or dyspraxia and/or MI syndrome

<table>
<thead>
<tr>
<th>Screening test</th>
<th>$R^2$ squared</th>
<th>Adjusted odds ratio (OR)</th>
<th>$p$ Value</th>
<th>95% Confidence interval for adjusted odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutor Screening</td>
<td>0.578</td>
<td>41.333</td>
<td>&lt;0.001**</td>
<td>10.032</td>
</tr>
<tr>
<td>Battery</td>
<td></td>
<td></td>
<td></td>
<td>Lower band</td>
</tr>
<tr>
<td>LADS</td>
<td>0.399</td>
<td>16.771</td>
<td>&lt;0.001**</td>
<td>4.815</td>
</tr>
</tbody>
</table>

**$p<0.01$, *$p<0.05$.**

The sensitivity and specificity for the TSB and LADS for any of the three disabilities is shown in Table 5. Table 5 also summarizes the performance of the two screening tests in terms of the number of students referred for assessment and the results, including false negatives, at screening.

The comparison between the TSB and LADS indicates that the TSB gave a stronger overall performance. LADS has a higher specificity, but a lower sensitivity, than the TSB, which means that it will miss a higher proportion of students with disabilities; in this case 12 of the sample, as against three missed by the TSB. However, the corollary is that the TSB resulted in eight assessments that proved negative, as against only four from LADS, and this has implications in terms of both cost and student stress levels.

It is worth noting that while it may have seemed at the outset that the relatively high sensitivity of the TSB might be due in part to the employment of a DASTA ARQ of 0.857 rather than 1.0, the three students who would have been excluded on the higher quotient of 7 (scoring exactly 6 on DASTA) all proved to be positive at assessment for at least two disabilities.

The next section reports the results of our exploration for a better performing screening for the three disabilities.

### An Enhanced Screening Battery

As discussed above, we need a screening test that is a good predictor for any and all of the three disabilities identified. This paper does not seek to delineate exact overlaps between the labels ‘dyslexic’ and ‘dyspraxic’.

In seeking to improve the screening process, we realized that it would be desirable to include the ‘gold standard’ test for M-I syndrome (the use of the Wilkins Rate of Reading Test® employing Cerium Overlays) as part of the screening process, rather than include it as part of the full assessment. Making this change in practice, and including the M-I syndrome test as the first part of a two-part screening process, would have ensured that this first stage immediately identified 51% (18/35) of those with a disability. As noted above, all but one of these students had another disability in addition to M-I syndrome. This means that presence of the syndrome is itself a powerful indicator of the other two disabilities.

Having identified all those students with M-I syndrome, we could employ the second part of the screening process to identify dyslexia and dyspraxia.
undertook a stepwise logistic regression on all the subtests included in Table 2, excluding *months forward*, and included LADS as a potential subtest, using data for the 56 students who screened negative for M-I syndrome. We excluded *months forward* because the data showed it to be a poor discriminator: only one student failed this subtest. The dependent variable was ‘having dyslexia and/or dyspraxia’. The $R^2$ squared for this model was 0.753, and the performance of the five subtests included in the model is summarized in Table 6.

Using the coefficients shown in Table 6, and the constant $-6.775$ generated by the stepwise regression, alternative sensitivity and specificity combinations for the New Screening Battery were calculated on the basis of different probability thresholds for part two of the screening process, as shown in Table 7. Table 7 also summarizes the performance of the New Screening Battery in terms of the number of students referred for assessment and the results, including false negatives, at screening.

Table 7 shows that a probability threshold of 0.06 for part two of the screening process allows the overall performance of the new screening process to achieve 100% sensitivity. However, the associated specificity is only 59%. The strongest performing combination of sensitivity and specificity is 94% and 92%, respectively, and this is achieved with a threshold of 0.50. This overall

Table 6. Stepwise logistic regression results: odds ratios and $p$ values of association for the subtests included in the model and dyslexia and/or dyspraxia

<table>
<thead>
<tr>
<th>Subtests</th>
<th>Coefficient</th>
<th>Adjusted odds ratio (OR)</th>
<th>$p$ Value</th>
<th>95% Confidence interval for adjusted odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower band</td>
</tr>
<tr>
<td>Polysyllables</td>
<td>4.068</td>
<td>58.433</td>
<td>0.018*</td>
<td>1.998</td>
</tr>
<tr>
<td>Familial incidence</td>
<td>2.947</td>
<td>19.051</td>
<td>0.061*</td>
<td>0.873</td>
</tr>
<tr>
<td>Digit span</td>
<td>1.638</td>
<td>5.145</td>
<td>0.014*</td>
<td>1.389</td>
</tr>
<tr>
<td>One minute writing</td>
<td>1.037</td>
<td>2.820</td>
<td>0.043*</td>
<td>1.033</td>
</tr>
<tr>
<td>LADS</td>
<td>3.916</td>
<td>50.178</td>
<td>0.006**</td>
<td>3.035</td>
</tr>
</tbody>
</table>

**$p<0.01$, *$p<0.05$.**

Table 7. Alternative sensitivity and specificity combinations for the New Screening Battery for dyslexia and/or dyspraxia and/or Meares-Irlen syndrome, and summary performance measures

<table>
<thead>
<tr>
<th>New screening battery (part two) probability threshold</th>
<th>New screening battery: performance of parts one and two combined ($n = 74$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (%)</td>
</tr>
<tr>
<td>0.06</td>
<td>100</td>
</tr>
<tr>
<td>0.10</td>
<td>97</td>
</tr>
<tr>
<td>0.50</td>
<td>94</td>
</tr>
</tbody>
</table>
performance is superior to the TSB, which achieved sensitivity of 91% and specificity of 79% (Table 5).

DISCUSSION

Limitations

While it would have improved the validity of the research to assess the 18 participants who screened negative in both screenings, and were assumed to be ‘true negatives’, this was financially prohibitive. However, the finding that all 10 students who volunteered for assessment, having screened negative using both tests, were found to be negative for all three target disabilities, broadly justifies our decision.

A second caveat is that caution must be exercised when applying the conclusions of our analysis to other student populations. Some of the characteristics of the students included in the analysis are similar to the wider UW population. For example, the age range of the students included in the analysis reflected that of the UW student population as a whole: the sample had 77% under 25 compared with 78% in UW, and 23% were aged 26–50, compared with 20% in this age range overall at UW, while none were over 50 compared with 2% at UW. We found a high incidence among those who had come from Access courses and again the proportions were similar, with 8% of the sample and 7% of UW students taking that route to university.

However, the incidence (47%, 35/74) in our sample of students identified with disabilities is much higher than that likely to be present in the wider student population. Volunteers, by definition, are not randomly chosen, so a higher incidence of disability was to be expected than in a random sample of the student population; those suspecting dyslexic tendencies in themselves were more likely to volunteer. We chose trainee teachers specifically in order to minimize this effect, as we expected that even those who had no suspicion that they might be dyslexic would have some professional interest in dyslexia. A questionnaire established that this was in fact one of the reasons for volunteering in 80% (29/36) of those who replied.

Hence, our sample is not likely to be representative of the wider student population in terms of level of disability. It is important to emphasize that because of this, the sensitivity and specificity of the screening tests could differ if applied to a random group of students (Sackett & Haynes, 2002). The characteristics of the sample used in our analysis are probably closer to those of the students actually being regularly screened at UW than to those of the UW student population as a whole, and a random sample would not have been possible within the practical constraints of this study.

Another possibility is that excluding 14 of our sample, because of incomplete data on M-I syndrome, might have distorted the dataset in favour of those who had a disability. Table 1 shows that this was not the case, but rather the opposite.
Tutor Screening Battery

The performance of each individual subtest’s association with the three disabilities was varied, as set out in Table 3. The logistic regression analysis identified five of the 15 subtests as being significantly associated with the three disabilities at below the 5% level: left/right confusions ($p = 0.019$), months reversed ($p = 0.042$) from BDTA, and one minute reading ($p = 0.011$), phonemic segmentation ($p = 0.022$) and two minute spelling ($p = 0.032$) from DASTA. Three other subtests were borderline: polysyllables ($p = 0.076$) (BDTA), digit span ($p = 0.082$) and nonsense passage reading ($p = 0.066$) (DASTA).

The $p$ values from these significant DAST subtests can be compared with the selection made for predicting dyslexia alone by Harrison and Nichols (2005) (one minute reading, phonemic segmentation, two minute spelling, nonsense reading and one minute writing) and Hatcher, Snowling, and Griffiths’ (2002) non-DAST results (nonword reading, spelling, digit span and writing speed). We can conclude that three individual subtests of DAST, one minute reading, phonemic segmentation and two minute spelling are effective at identifying dyslexia and/or dyspraxia and/or M-I syndrome as well as dyslexia alone.

TSB Versus LADS

Looking at sensitivity first, the TSB performs well, at 93%, compared with LADS at 66% (Table 5). Conversely, LADS provides best specificity: 90% compared to TSB’s 79%. In practice, if the results of this study hold good for a wider population, this would mean that for every 100 students screened using the TSB, 54 would be referred for assessment. Of these, 43 would be assessed as positive and 12 as negative. Four students with a disability would be missed (false negatives). In comparison, LADS would refer only 36 for assessment, resulting in fewer negative assessments (only five as opposed to 12 from the TSB), but 16 students with a disability would be missed.

As discussed above, acceptable levels of sensitivity and specificity vary with the availability of treatment/support and the cost of conditions remaining unidentified (Grimes & Shultz, 2002; Streiner, 2003). In this case, support is available for those who are assessed positively. Thus, LADS’ relatively poor sensitivity, turning away a comparatively high proportion of students who in fact have a learning-related disability, may be seen to have an unacceptably high cost in human terms.

Improving Practice

Having established that the TSB is a good instrument for predicting the targeted disabilities in HE students, we next considered whether this battery could be shortened, without significant loss of accuracy. The analysis reported above shows that this is indeed possible. In particular, the co-morbidity of M-I syndrome with dyslexia and dyspraxia illustrated in Figure 2 shows that M-I syndrome is itself a valid predictor of dyslexia and dyspraxia. It is so highly correlated with dyslexia and dyspraxia ($p<0.001$, $r = 0.60$) that any student who is identified with M-I syndrome can be immediately referred to assessment on
that basis alone; as stated above, in our sample this would have identified 51% of those with a disability.

This finding supports our desire to administer tests for M-I syndrome during the screening process, rather than waiting for the assessment stage. Assessing M-I syndrome during screening for dyslexia and dyspraxia allows direct referral from screening to a specialist optometrist. This change in procedure would also mean that while awaiting assessment students could familiarize themselves with any preferred overlay, and that its use during assessment would allow clearer identification of which difficulties might be assignable solely to M-I syndrome, and which must be attributed to an SpLD.

Where M-I syndrome is not present, a new selection of subtests can be employed to predict dyslexia and dyspraxia, which leads to a better performance than either the original TSB or LADS ($R$ squared is 0.753 compared to 0.578 and 0.399, respectively). This selection consists of LADS itself, plus the subtests polysyllables, familial incidence, digit span and writing speed.

This result is consistent with our finding that while the LADS specificity is high, at 90%, its sensitivity is too low, at 66%. The addition of these four extra subtests maintains its specificity while markedly increasing sensitivity. Using optimal cutoffs (discussed above, see Table 7), this combination gives a sensitivity of 94% and specificity of 92%.

If we return to the list of deficits that our screening aims to identify, this new set of tests covers all the items. Digit span assesses working memory, familial incidence points towards any genetic factors, while polysyllables measures phonological processing and may also pick up dyspraxic students who have underlying problems of articulation. Writing speed will also identify difficulties of both phonological and visual processing, and further may provide some measure of hand–eye coordination. LADS encompasses working memory, phonological and visual processing, sequencing and orientation, spelling, reading, and again some assessment of hand–eye coordination in the use of the keyboard. Clearly the M-I syndrome test covers general visual processing to some extent, as well as identifying levels of the syndrome itself.

It is interesting to note that stepwise logistic regression analysis has selected digit span from the range of subtests, since LADS itself contains a digit span test, its memory subtest. While the correlation between digit span and memory is strong ($p<0.01$, Pearson correlation 0.575) these are clearly not duplicate tests (which would generate a Pearson correlation of 1). This tallies with our experience, as many students reported that they found the LADS memory test easier than the digit span because in the former they were typing memorized numbers into a keypad, which itself served as both a visual and a kinaesthetic memory aid, while in the latter they had no such assistance.

CONCLUSIONS

This study has confirmed that the TSB performs more strongly than LADS and that it performs well. Nevertheless, tutor screening is comparatively expensive. The investigation into improving this battery has produced a different approach to screening. The benefit of this is two-fold: the change in practice enables improved support to students in terms of more rapid identification of M-I
syndrome, and the improved selection of subtests brings better performance without increasing the screening time required.

The resource implications of this lie beyond the scope of this paper and warrant further investigation. To the extent that there is a trade-off between the superior performance of tutor screening and its relatively high cost compared with computerized screening, the new screening battery goes some way towards a compromise. Moreover, while still achieving a high level of sensitivity, it should result in fewer false positives, thus cutting down the number of assessments to be made, given the caveats about differing populations above. The research thereby provides insight relevant to policy decisions in HE about screening methods and performance.

ACKNOWLEDGEMENTS

This research could not have been conducted without the support of the Centre for Inclusive Learning Support at the University of Worcester, and its Director, Dr Val Chapman, NTF. The authors would also like to thank all the following: Julian Brown, Louise Smith and Fran Summerfield, members of the project team; LUCID research for allowing us special rates on the LADS programme; the administrative staff of the Disability and Dyslexia Service, University of Worcester; and all those students who participated.

References


