

Consensus or contradiction? A review of the current research into the impact of granting extra time in exams to students with Specific Learning Difficulties (SpLD)

Helen Duncan¹ and Catherine Purcell²

¹Senior Neurodiversity Adviser. University of Cambridge, Keynes House, Cambridge, CB2 1QA

²Senior Lecturer, School of Healthcare Sciences, College of Biomedical and Life Sciences, Cardiff University, Heath Park, Cardiff. CF14 4XN

Abstract

Timed examinations continue to be a common educational assessment method employed to evaluate a student's knowledge, ability and skills in their subject area in further and higher education (Zuriff, 2000; Lovett, 2011). Extra time is the most common adjustment that students with specific learning difficulties (SpLD) are granted in these exams. This adjustment aims to provide students with SpLD parity with their typically developing (TD) peers in exams by ensuring that the differences in scores between candidates reflects the differences in their subject ability, not in their speed (Bolt, Decker, Lloyd & Morlock (2011); Lovett (2011)). Numerous studies have been conducted with the aim of identifying whether extra time confers an advantage or simply enables students with SpLD to more accurately demonstrate their potential in exam situations. This paper summarises the existing empirical research investigating the practice of granting this exam adjustment in order to identify what the current body of research concludes is the effect of receiving extended time on the results of candidates with SpLD. This investigation found no consensus of opinion about the fairness of this adjustment (i.e. whether it is only students with SpLD who benefit or whether all students benefit similarly from additional time). However, despite the inconsistencies in findings between studies a pattern does emerge across all the studies as a whole, which suggests that the differences in outcomes may be explained by differences in the appropriateness of the time limits imposed by each study.

1. Introduction

Timed examinations continue to be a common educational assessment method employed to evaluate a student's knowledge, ability and skills in their subject area. Extra time is the most common adjustment that students with Specific Learning Difficulties (SpLD) are granted in exams (Zuriff, 2000; Lovett, 2011). This adjustment aims to provide students with SpLD parity with their typically developing (TD) peers in exams by ensuring that the differences in scores between candidates reflects the differences in their subject ability and are not the result of their disability (Bolt, Decker, Lloyd & Morlock, 2011; Lovett, 2011). At the same time, extra time in exams is not intended to confer an advantage (Sireci & Parker 2005). Any improvement in performance by the SpLD candidate in receipt of extra time should thus represent a

more accurate score that is a valid representation of the student’s subject knowledge and skills. It should not simply result in an increase in marks that represents an over-inflated score as this would challenge the validity of the exam result (Lewandowski, Cohen, & Lovett, 2013).

The granting of extra time in exams to students with SpLD is contentious (Lovett, 2010). Defenders of the practice argue that awarding extra time in exams simply removes any disadvantage to a candidate with SpLD that arises as a direct result of the time constraints of the exam, as candidates with SpLD take longer to demonstrate their subject knowledge in writing (McKimm, 2012; Singleton, 1999; Haladyna & Downing, 2004). However, critics of the practice argue that awarding additional time in exams allows students with SpLD to produce longer answers, include more details, and produce a richer response that more fully answers the question. This results in over-inflated marks (Zuriff, 2000). The aim of this paper is to review the range of published evidence that explores whether granting extra time to students with SpLD confers an advantage in exams or simply levels the playing field.

2. Search strategy

In order to review the published studies exploring the impact of granting extra time in exams to students with SpLD, a search across a range of educational and multidisciplinary databases for conference and peer-reviewed journal papers was carried out (Booth, Sutton & Papaioannou, 2016). The key terms used for this search included, ‘exam’ and ‘adjustments’, or ‘accommodations’, ‘arrangements’, ‘access’, ‘special arrangements’, ‘test accommodations’, extra time’, ‘additional time, ‘timed tests’. This core search was combined with SpLD specific terms including, ‘specific learning difficulties’, learning disabilities’, ‘reading disabilities’, ‘dyslexia’, ‘dyspraxia’, ‘DCD’. The databases searched are shown in Table 1.

Table 1: Databases of Journals and conference papers searched

Database	Coverage
British education index (including conference abstracts)	Education (UK & Europe)
Australian education index	Education (Australasia)
Education resources Information Centre (ERIC)	Education (US)
PubMed	Multidisciplinary
Scopus	Multidisciplinary
Google Scholar	Multidisciplinary
Web of Science (including Social Sciences and humanities conference data base)	Multidisciplinary
iDiscover (searches Cambridge University library collection of printed and electronic texts, journals, databases, e-journals and on-line resources)	Multidisciplinary
Education Research: Higher Education Academy	Education
PsychINFO, PsychARTICLES, PsycBOOKS	Psychology and related disciplines

3. Inclusion and exclusion criteria

The search included studies published between 1980 and 2018. Peer-reviewed papers that used an experimental or quasi-experimental design were included, as were articles that involved systematic reviews and meta-analyses. Articles that used only qualitative approaches or were not peer reviewed were excluded, as were studies with less than 80% power to detect an effect.

The term ‘specific learning difficulty’ can be wide-ranging and includes a variety of characteristics (Frederickson & Reason, 1995). As the purpose of this paper is to explore the effect of extra time in exams, the main characteristics associated with SpLD that are the focus of this paper are those that are pertinent in the context of a timed, written, examination. For this reason, the term ‘specific learning difficulty’ here is confined to diagnoses that pivot mainly on processing difficulties. Studies investigating the impact of exam arrangements on students with other disabilities, including those with Attention Deficit Hyperactivity Disorder or Autism Spectrum Disorder have been excluded.

4. Overview of the survey

A survey of the literature identified 16 studies that suggested that the granting of extra time did not confer any advantage to students with SpLD (sample sizes between 32 and 79,963 participants). These are shown in Table 2.

Table 2: Studies that found extra time did not confer an advantage

Author and Source	Title of study
Alster, E. H. (1997). <i>Journal of Learning Disabilities</i> . 30 pp 222-227	The effects of extended time on algebra test scores for college students with and without learning disabilities
Cahalan-Laitusis, C., King, T. C., Cline, F., & Bridgeman, B. (2006). Research Report 2006-4 New York, NY: College Board.	Observational timing study on the SAT Reasoning test for test-takers with learning disabilities and/or ADHD.
Cohen, A., Gregg, N., & Deng, M. (2005). <i>Learning Disabilities Research & Practice</i> . 20 (4) pp225-233	The Role of Extended Time and Item Content on a High-Stakes Mathematics Test
Crawford, L. Helwig, R., Tindal, G. (2004). <i>Journal of Learning Disabilities</i> . 37 (2) pp132-42	Writing performance assessments: how important is extended time?
Duncan, H., & Purcell, C. (2015) <i>Journal of Widening Participation and Lifelong Learning</i> . 19 (2) pp 6 – 26)	Equity or Advantage? The effect of receiving access arrangements in university exams on Humanities students with Specific Learning Difficulties (SpLD)
Hill, G.A. (1984). Doctoral dissertation, Texas Technical University, Lubbock	Learning disabled college students: The assessment of academic aptitude.
Finch, H., Barton, K., & Meyer, P. (2009). <i>Educational Assessment</i> . 14(1) pp38-56	Differential item functioning analysis for accommodated versus non-accommodated students

Huesman, R. L., & Frisbie, D. A. (2000). Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA	The validity of ITBS reading comprehension test scores for learning disabled and non-learning disabled students under extended time conditions'
Lindstrom, J. H., & Gregg, N. (2007). <i>Learning Disabilities Research & Practice.</i> 22 (2). pp85-95	The role of extended time on the SAT® for students with learning disabilities.
Lindstrom, J.H. (2010). <i>Intervention in School and Clinic.</i> 46 (5). pp 5-12.	Mathematics Assessment Accommodations: Implications of Differential Boost for Students with Learning Difficulties.
Lesaux, N.K., Pearson, M.R., & Siegel, L.S. (2006). <i>Reading and Writing.</i> 19 (1). pp21-48	The effects of timed and untimed testing conditions on the reading comprehension of adults with reading disabilities
Ofiesh, N., Mather, N., & Russell, A.(2005) <i>Journal of Psychoeducational Assessment.</i> 23 pp35-52	Using speeded cognitive, reading, and academic measures to determine the need for extended test time among university students with learning disabilities
Runyan, M. K. (1991) <i>Journal of Learning Disabilities.</i> 24 (2) pp104-8	The Effect of Extra Time on Reading Comprehension Scores for University Students With and Without Learning Disabilities
Runyan, M. K., & Smith, J. (1991). <i>Journal of Legal Education.</i> 41. pp 317–349.	Identifying and accommodating learning disabled law school students.
Weaver, S.M. (1993). Doctoral dissertation, University of Toronto, Canada	The validity of the use of extended and untimed testing for postsecondary students with learning disabilities.
Ziomek, R.L., & Andrews, K.M. (1998). American College Testing Program. <i>ACT Research Report Series.</i> Iowa City. IA	ACT Assessment Score Gains of Special-Tested Students Who Tested at Least Twice

At the same time, 12 studies were identified that suggested that the granting of extra time over-inflated the scores of all candidates (sample sizes between 30 and 119,490 participants). These are shown in Table 3.

Table 3: Studies that suggest extra time confers an advantage

Author and Source	Title of study
Amodeo, A., Marcus, L.A., Thornton, A.E., Pashley, P.J. (2009). Law School Admission Council. LSAT Technical Report 09-01. <i>LSAC Research Report Series.</i>	Predictive Validity of Accommodated LSAT Scores for the 2002-2006 Entering Law School Classes.
Cahalan, C., Mandinach, E.B., & Camera, W. J. (2002). The College Board. Research Report No. 2002-5. College Entrance Examination Board. New York	Predictive Validity of SAT I Reasoning Test for Test-Takers with Learning Disabilities and Extended Time Accommodations
Elliott, S. N., Marquart, A.M. (2004). <i>Exceptional Children.</i> 70 (3). pp349 - 367	Extended time as a testing accommodation: Its effects and perceived consequences.
Halla, J. W. (1988). Unpublished doctoral dissertation, Texans Tech University, Lubbock.	A psychological study of psychometric differences in Graduate Record Examination General Test scores between learning disabled and non-learning disabled adults.

Lewandowski, L. J., Lovett, B.J., & Rogers, C.L. (2008). <i>Journal of Psychoeducational Assessment</i> 26 (4). pp 315- 324	Extended Time as a Testing Accommodation for Students with Reading Disabilities: Does a Rising Tie Lift All Ships?
Lewandowski, L. J., Lovett, B.J., & Rogers, C.L. (2008). <i>Journal of Psychoeducational Assessment</i> , 31(3). pp 326-336.	Effects of Extended Time Allotments on Reading Comprehension Performance of College Students With and Without Learning Disabilities.
Lovett, B.J. (2011). <i>Communique: The newspaper of the National Association of School Psychologists (NASP)</i> . 39 (1)	Extended Time Accommodations: What does the Research say?
Lovett, B. J. (2010). <i>Review of Educational Research</i> . 80 (4). pp 611 - 638	Extended Time Testing Accommodations for Students with Disabilities: Answers to Five Fundamental Questions
Mandinach, E.B., Bridgeman, B., Cahalan-Laitusis, C., & Trapani, C. (2005). College Board Research Report 2005-8. New York. NY: College Board	The impact of extended time on SAT test performance.
Thornton, A.E., Reese, L.M., Pashley, P.J., & Delassandro, S.P. (2001). Law School Admission Council, Newtown, PA	Predictive Validity of Accommodated LSAT Scores. Technical Report. LSAC Research Report Series.
Zurcher, R., & Bryant, D.P. (2001). <i>Journal of Learning Disabilities</i> 34(5). pp 462-471	The Validity and Comparability of Entrance Examination Scores After Accommodations Are Made for Students with LD
Zuriff, G.E. (2000). <i>Applied Measurement in Education</i> . 13. pp 99-117	Extra examination time for students with learning disabilities: An examination of the maximum potential thesis.

In addition, 4 systematic reviews were identified that concluded that the body of research evidence yields no consensus. These are shown in Table 4.

Table 4: Studies that identified no overall consensus

Author and Source	Title of study
Bolt, S. E. (2004). Paper presented at the annual conference of the National Council on Measurement in Education, San Diego, CA	Using DIF analyses to examine several commonly held beliefs about testing accommodations for students with disabilities.
Gregg, N., & Nelson, J.M. (2010). <i>Journal of Learning Disabilities</i> . 45. pp 128 - 38	Meta-analysis on the Effectiveness of Extra Time as a Test Accommodation for Transitioning Adolescents with Learning Disabilities: More Questions than Answers.
National Centre on Educational Outcomes (2000- 2016). NCEO APR Snapshot Briefs. 11. National Centre on Educational Outcomes (NCEO).	Assessment Accommodations use by Students Receiving Special Education Services.
Sireci, S. G., Scaroati, S. E., & Li, S. (2005). <i>Review of Educational Research</i> . 75 (4). pp 457-490	Test accommodations for students with disabilities: An analysis of the interaction hypothesis.

5. Maximum Potential Thesis (MPT)

Of the 16 identified studies that suggest that granting extra time fails to confer an advantage, 6 found that additional time improved the exam performance of students with SpLD only, with TD participants failing to improve their marks with extra time (Cohen, Gregg & Deng, 2005; Hill, 1984; Huesman & Frisbie, 2000; Lesaux, Pearson & Siegal, 2006; Runyan 1991). These 6 studies posit that only students with SpLD improve their performance when granted additional time in exams as TD students are able to reach their potential under standard timed testing conditions. This is known as the Maximum Potential Thesis (MPT) and pivots on the assumption that candidates are unable to perform better than the limit of their knowledge and skills in the exam simply because they are given more time. Students with SpLD, however, process information more slowly and so standard exam conditions may not be sufficient for these candidates to show the extent of their knowledge and ability without extra time (Bolt, 2004).

Runyan (1991), for example, compared the performance in a written comprehension test between 16 University students with SpLD and 16 TD peers, with both groups taking the test under standard time conditions and also with additional time. Runyan found that in the 20 minute timed comprehension task, the scores of TD students significantly exceeded those of matched participants with SpLD when both groups were assessed under standard time conditions. However, the scores of the students with SpLD improved materially when awarded additional time to the extent that the gap in performance was closed. When comparing the scores of the participants with SpLD in receipt of additional time to the scores of the TD participants who were tested under standard timed conditions, no statistically significant differences were identified and there were no significant differences in scores between the two groups when both groups were tested under extended time conditions.

Furthermore, Runyan (1991) found that the TD students not only completed the comprehension task within the 20 minutes normally allocated but also failed to show significant improvements in their scores when granted additional time. Runyan argues that the TD participants were already achieving their 'maximum potential' within the standard time allotted and hence did not improve their scores when allowed more time. Conversely, the participants with SpLD failed to complete the comprehension task within the 20 minutes (albeit they answered those questions they attempted correctly). Runyan concludes that the additional time simply levelled the playing field and normalised the performance of the participants with SpLD to that of their TD peers. The extra time only benefitted those participants who needed it due to their disability.

There are, however, methodological weaknesses in this research study that may undermine the reliability of the findings. Firstly, participants were advised that they were taking a timed test but were not informed of the amount of time available in which to complete it (they were simply advised to work as quickly and accurately as possible). When granted extra time, participants were not permitted to change or review the answers that they had made to the test during the first 20 minutes (as these first 20 minutes represented the 'timed' conditions). Thus, it could be argued that those participants who worked quickly and completed the test within the standard time were unable to improve their marks when given additional time, even if they had not performed to their potential, as they were not permitted to revisit or revise any answers already given (Zuriff, 2000). This may have resulted in a 'ceiling effect' for these participants (most likely to be the TD participants) and so could explain why granting

additional time to students with SpLD reduced the gap in performance between SpLD and TD participants in this study.

6. Differential Boost hypothesis.

In the study by Runyan (1991), only the participants with SpLD benefitted from extra time. The TD participants did not improve their scores. Other studies cast doubt on these findings and conclude that all students show performance improvements when granted extra time, and students with SpLD show significantly greater gains, or a 'differential boost' in comparison with their TD peers (Alster, 1997; Lindstrom & Gregg, 2007; Lesaux, Pearson & Siegal, 2006; Ofiesh, Mather & Russell, 2005). This 'differential boost hypothesis' predicts that there will be greater differences in performance between SpLD and TD candidates under standard time conditions than under additional time, because candidates with SpLD will 'close the gap' when granted extra time as students with SpLD benefit more from the additional time than their TD peers (Sireci, Scarpati & Li, 2005).

Lesaux, Pearson & Siegel (2006), for example, investigated the differential boost hypothesis by assessing performance in a reading comprehension test between 22 adult participants with SpLD and 42 TD peers. Both groups were tested in both timed and extended time conditions. All participants were matched for age and IQ, but there was a wide range of socio-economic and educational backgrounds, with the SpLD group having spent fewer years in education (mean = 11.95 years; SD 2.36) than the TD controls (mean = 13.95 years; SD 2.91). All participants undertook the Nelson Denny Reading Comprehension Test (Brown, Fishco & Hanna, 1993), comprising 7 passages of prose reading followed by 38 multiple-choice questions. Participants were given the first form of the test under the standard timed conditions of 20 minutes and were then given 100% extended time (40 minutes) in which to complete the parallel form of the test. The outcomes were compared to identify any between-group and within-group differences in performance when taking the test under standard time and additional time conditions. Comparisons were made between the number of questions attempted, the number of correct responses achieved, and the number of errors made under both sets of time conditions.

When taking the test under standard time conditions, the TD group attempted significantly more questions on average (34.90 questions out of the possible 38) than the SpLD group (21.64 questions out of 38). In addition, the TD group achieved significantly higher numbers of correct responses; accurately answering an average of 30.40 questions in comparison to the SpLD group who achieved an average of 12.68 correct responses. When granted 100% additional time, both groups answered all of the questions, although the TD group still achieved a significantly higher number of correct answers than the SpLD group. The mean number of correct answers achieved by the TD group increased by 3.75 to 34.15 while the mean number of correct answers achieved by the SpLD group increased by 10.91 to 23.59. Thus, the group with SpLD improved their performance with extra time such that the difference in performance between the two groups, while still evident, reduced significantly. Although the TD group also improved their performance in extended time conditions in comparison to standard time conditions, the group with SpLD showed a much greater improvement and benefitted more from the additional time than the TD control group (a 'differential boost').

However, even with extra time, the group with SpLD still failed to perform at the same level as their TD peers. This may be due to the statistically significant differences in the language, word reading and working memory skills between the two groups coupled with the fact that the SpLD group did not have age appropriate reading ability. The group with SpLD in this study also had significantly fewer years of education than the TD control group. A proportion of the SpLD participants had failed to complete compulsory education, with very few continuing to post-secondary education, and so may have had less exposure to structured teaching of reading prose and comprehension skills and also reduced opportunities to practice these skills. These differences between the two groups (especially the fact that the SpLD group had only basic literacy skills) could account for some differences in performance in a reading comprehension test and the variance in scores.

Nonetheless, the overall findings of the study by Lesaux, Pearson & Siegel (2006) showed that participants with SpLD benefitted from additional time, significantly improving their scores and being able to complete the test. While the TD group improved their performance with extra time this improvement was not statistically significant and this group was already able to perform within the average range under standard time conditions.

7. Extra time over-inflates scores

Studies that show that candidates with SpLD benefit more from extra time than their TD peers are contested by studies that conclude that extra time benefits all candidates, over-inflating marks and undermining the validity of the exam results (Amodeo, Marcus, Thornton & Pashley, 2009; Cahalan, Mandinach & Camara, 2002; Elliott & Marquart, 2004; Halla, 1988; Lewandowski, Lovett, Parolin, Gordon & Coddling, 2007; Lewandowski, Lovett & Rogers, 2008; Thornton, Reese, Pashley & Dalessandro, 2002; Zurcher & Bryant, 2001).

Lewandowski, Cohen & Lovett (2013), for example, examined the effect of granting 50% and 100% extra time to SpLD and TD students in reading comprehension tests to see whether there was a greater between-group difference in performance under standard time than with extra time. Lewandowski, Cohen & Lovett posited that, if the granting of extra time is justified, TD participants should achieve higher marks than participants with SpLD when both groups take the test under standard time conditions. However, when both groups are granted extra time there should be no between-group differences in performance. TD candidates should not increase their scores with additional time as it is expected that they reach their full potential under standard time conditions. This study comprised 26 University undergraduate students with SpLD and 50 TD peers, matched for gender, ethnicity and age. The two parallel reading comprehension forms of the Nelson Denny Reading Comprehension Test (Brown, Fishco & Hanna, 1993) were combined to produce a single test consisting of 14 passages to read and 76 multiple choice questions to answer. This aimed to circumvent any possible ceiling effect that may arise when extra time was implemented. Each participant's test was marked after 15 minutes (i.e. standard time), again after 22.5 minutes (i.e. with 50% extra time) and a third time after 30 minutes (i.e. with 100% extra time). The numbers of correct responses for each group, as well as the number of questions attempted and the percentage of correct responses, were compared. The

results showed that, although neither group was more accurate in their response than the other, the TD participants scored increasingly higher than the SpLD group as time allocation increased. This contradicts the differential boost hypothesis that participants with SpLD will ‘catch up’ with their TD peers when given additional time. Instead, Lewandowski, Cohen & Lovett found that all participants improved their scores with extra time, but that TD participants showed greater gains than the participants with SpLD.

However, when the scores of the participants with SpLD at standard time (15 minutes), 50% extra time (22.5 minutes), and 100% extra time (30 minutes) were compared with the scores achieved by the TD peers under standard time (i.e. their scores after 15 mins), the performance of the SpLD group was lower than the TD group after standard time (15 minutes) but higher than the scores of the TD group after 50% extra time and also after 100% extra time. This showed that the SpLD group performed below the TD group under standard conditions but outperformed the TD group when granted extra time in comparison to the TD group performance under standard time conditions. Lewandowski, Cohen & Lovett suggest that this indicates that extra time confers an advantage (and overestimates the ability of the SpLD participants) as the SpLD group should not achieve significantly higher scores when given extra time than the control group achieved under standard time, but that the scores for each group should be equalised when the group with SpLD are given extra time.

Lewandowski, Cohen & Lovett (2013) argue that these findings show that all students improve their performance when they receive additional time, not only students with SpLD. The study also found that TD students benefitted more from the extra time than those with SpLD, with the gap between the performance of the participants with SpLD and their TD peers widening as time increased. In addition, the finding that the candidates with SpLD answered more questions and achieved greater numbers of correct responses when given extra time than the TD peers achieved under standard time is presented as evidence that extra time “*goes beyond leveling the playing field, and may reverse the playing field in favour of students with SpLD*” (Lewandowski, Cohen & Lovett, 2013, p333). This calls into question the validity and appropriateness of allowing extra time as an exam access arrangement.

However, caution should be exercised when generalising the findings of this study to the population as a whole due to methodological problems that may undermine the reliability of these findings. The Nelson Denny Reading Comprehension subtest was designed (and standardised) as a 20-minute test. Therefore, by doubling the number of test items (through combining the two parallel forms) while restricting the ‘standard time’ to 15 minutes means that the test is not deemed fully achievable for either SpLD or TD participants within the ‘standard time’ (as the standard time that the original designers had allocated to this combined activity was 40 minutes). Lewandowski, Cohen & Lovett argue that this controls for the ‘ceiling effect’ as the participants will not have reached the ceiling of their ability (as determined by the number of correct answers) within the time permitted but can continue to answer further test items (i.e. time will run out before the number of answers are completed). However, this removal of the test ceiling may account for the finding that TD participants improved their performance when given extra time, as even double extra time (30 minutes) still represented less than the standard time that the test deemed necessary for the completion of both forms concurrently (standard time for both tests concurrently is 40

minutes). Given that the participants with SpLD performed below their TD peers in this study when allowed the 'standard time' of 15 minutes, it is not surprising that they performed below their TD peers when both groups were given 'additional time' of 50% as this was still insufficient time for either group to complete the task. Even when given 100% extra time (30 minutes), this still represented 25% less than the standard time normally allowed by the Nelson Denny test itself (i.e. 40 minutes for 76 questions). When granted 30 minutes to complete the test (i.e. 100% extra time for the purposes of this study) the TD participants completed more of the questions than the SpLD participants. That is to say, the TD group completed, on average, 80% of the test items (SD 9.6) in 30 minutes, while the SpLD group completed, on average, 64% of the test items (SD 12.49) in 30 minutes. This suggests that the TD group should be able to complete all 76 items on the test within the normal standard time of 40 minutes that the Nelson Denny Comprehension test itself sets (given that the TD group had completed 80% of the test items in 75% of the standard time allowed by the test). By contrast, the SpLD group would require extra time to finish the test (as the SpLD group had completed 64% of the test items in 75% of the standard time allowed by the test). It is worth noting that there were no between-group differences in accuracy of response, which suggests that the SpLD participants were slower (but not less accurate) in answering the questions.

Thus, the conclusion of Lewandowski, Cohen & Lovett (2013) that additional time confers an advantage to all students who receive it may only be the case where an exam is not achievable for any candidate in the time allowed. Given the unrealistic time limits placed on the participants in this study, rather than providing evidence to suggest that extra time is not an appropriate adjustment as it improves the scores of all students, not just disabled students, it could be argued that this study provides evidence to suggest that exam designers should calculate the standard exam time carefully to ensure that there is sufficient time for candidates to complete the exam.

8. Systematic review

Given that the body of research into the granting of exam access arrangements to students with SpLD appears to be characterised by contradictory findings, Sireci, Scarpati & Li, (2005) conducted a systematic review of eight research studies into the impact of additional time on exams. This aimed to statistically analyse, synthesise and summarise the findings and data from across the range of related studies with the aim of resolving the debate.

Of these 8 studies, 5 were experimental (Elliott and Marquart, 2004; Runyan, 1991; Halla, 1988; Hill, 1984; Weaver, 1993) and 3 were quasi-experimental (Huesman and Frisbie, 2000; Alster, 1997; Zurcher and Bryant, 2001). Three of the studies examined the impact of extra time on students with SpLD using the Nelson Denny Reading Comprehension Test (Brown, Fishco & Hanna, 1993) (sample sizes ranged from 30 to 526), two studies used generalised maths tests (sample sizes ranged from 31 to 124) and three studies used data from national tests, such as the Graduate Record Examination (GRE), the Scholastic Aptitude Test (SAT) or American College Testing (ACT) (sample size ranged from 34,012 to 52,667). Sireci, Scarpati & Li, (2005) concluded that some studies showed that just the SpLD group benefitted from additional time, some showed that while both groups increased their results, the

increase was greater for the SpLD participants than their non-disabled comparators and some showed that both groups benefitted equally. Sireci, Scarpati & Li also identified a range of limitations in the studies, namely that the Nelson-Denny Comprehension Test is a speeded test and thus “*all students would be expected to have score gains when extended time is given on a speeded test*” (Sireci, Scarpati & Li, 2005, p466). Overall, Sireci, Scarpati & Li concluded that while no clear consensus exists across studies, the outcomes suggest, on balance, that students with SpLD benefit more from additional time than TD students.

Gregg and Nelson (2012) similarly explored the impact of granting extra time in exams to students with SpLD through a meta-analysis that combined and evaluated the findings of 9 studies that had been conducted between 1986 and 2006 (Alster, 1997; Braun, Ragosta & Kaplan, 1986; Cahalan, Mandinach & Camara, 2002; Camara, Copeland & Rothschild, 1998; Cohen, Gregg & Deng, 2005; Lesaux, Pearson & Siegal, 2006; Lindstrom & Gregg, 2007; Ofiesh, Mather & Russell, 2005; Ragosta, Braun & Kaplan, 1991). This meta-analysis compared the test scores of participants with SpLD who were granted extra time with the test scores of TD participants who took the same test under standard conditions across all nine studies. The combined findings of the 9 studies indicated that the TD participants achieved higher scores than the participants with SpLD, despite the latter receiving extra time (mean effect size of $d = -.41$). In addition, the TD participants achieved higher scores than the participants with SpLD when both groups were granted extra time (mean effect size of $d = -.69$). Finally, when both groups took the test under standard conditions the TD participants achieved higher scores than the participants with SpLD (mean effect size of $d = -.86$).

Gregg and Nelson (2012) also made two within-group comparisons. Firstly, the test scores that the SpLD group achieved when granted extra time were compared with the scores they achieved when taking the test under standard conditions. This showed that the SpLD participants achieved higher scores when granted extra time (mean effect size of $d = .90$). Secondly, the test scores of TD participants taking the test under standard time were compared with their scores when granted extra time. This showed that the TD participants achieved higher scores when granted extra time (mean effect size of $d = .66$).

Overall, Gregg and Nelson (2012) concluded that participants with SpLD underperform in timed tests in comparison to their TD peers. Although this effect was reduced when participants with SpLD were permitted extra time, the extra time did not fully level the playing field or fully compensate for the effect of the disability on exam performance:

“The most significant finding of our meta-analysis is that...students with LD still underperform academically as compared to their normally achieving peers whether provided extended time or not on standardized tests. While students with LD perform significantly better when provided extended time, the accommodation does not erase the disability” (p136)

Although there are advantages of using a meta-analysis rather than simple narrative approach to a systematic review, it should be noted that the findings of a meta-analysis can be limited by the difficulties of comparing heterogeneous studies that use a wide variety of different designs and methodologies as well as by the inclusion of a number of studies that have a small sample size, as combining a number of small studies is not

the same as a single study with a large sample size. In addition, the inclusion of studies with weak internal, external or construct validity results in the meta-analysis combining studies of different qualities - thereby combining potentially reliable data with potentially unreliable data (Bryman, 2009).

In view of these limitations, Gregg and Nelson (2012) concluded that it would be problematic to generalise the findings of current studies in the field to the population as a whole, due to methodological limitations of the studies involved (especially the lack of ecological validity). This is an observation that is reiterated by the National Centre on Educational Outcomes (NCEO), which conducted a systematic review of 72 studies into the effects of exam access arrangements on students with disabilities between 1999 and 2015. The NCEO identified inconsistency in findings across the studies as well as a '*lack of authenticity of the testing circumstances*' (NCEO, 2015, p85). The lack of consensus in outcomes as well as poor ecological validity of the extant research identified by NCEO, highlights the need for research into the impact of exam arrangements in the context of the exam itself.

9. Limitations of current research

9.1 Ecological validity

The findings of studies investigating the impact of granting additional time to students with SpLD in timed tests are used to inform the debate around permitting extra time in 'real-life', high-stakes, exam situations. However, many of the research studies explored in this literature review adopt methodologies that raise questions relating to ecological validity. For example, most of these studies investigate the impact of granting extra time in reading comprehension or maths tests under standard and extended time conditions. However, the extent to which the findings of these studies can be transferred to the context of a school or university exam is questionable, given that many of the key processes involved in completing a reading comprehension or maths test differ significantly from those required for undertaking a formal exam. An examination testing the candidate's knowledge, understanding and skills in their taught subject area involves the recall of information and the ability to develop this into a novel argument. The candidate is required to generate and spell expressive vocabulary, drawing heavily on word retrieval skills. At the same time, presenting a cogent argument in their writing requires the planning, organising and editing of the candidate's ideas. By contrast, a reading comprehension test assesses the candidate's ability to understand a written passage and respond to questions based on the information (either explicit or implied) contained in the passage. Additionally, in a reading comprehension test the vocabulary required is largely given. The candidate is not generally required to generate new information, draw on recall or the interpretation of previously learned concepts, but instead the candidate is required to move or interpret information that is already given in a text to which they can continually refer. Fewer burdens are placed on the candidate's memory and recall, word retrieval, expressive language, planning, structuring and organisational skills; areas that tend to be vulnerable in students with SpLD (Reid, 2009).

Similarly, it can be argued that many of the processes involved in completing a maths test differ significantly from those required for undertaking a formal exam. For example, the maths tests used in the research studies cited above were not testing the

participants' knowledge of a recently taught course or curriculum. These were generalised maths tests that assessed a broad range of pre-existing maths knowledge and ability. The participants may have had different levels of knowledge of maths concepts and exposure to maths courses prior to the test (this potential variability was not considered by the researchers). By contrast, school or university exams aim to test the subject matter that has been studied through the year (or course) and assume that the participants have had comparable opportunities to be exposed to the body of knowledge being tested.

A study by Duncan & Purcell (2017), however, was conducted within the 'real-life' context of the mandatory high-stakes timed University exams. This quasi-experimental study examined existing data from the exam scripts of end of year university exams and compared the performance of 69 students with SpLD who were granted exam arrangements with the performance of 70 TD peers who took the same exam under standard exam conditions. The main areas of exam performance scrutinised were the length of exam paper (to ascertain whether the SpLD candidates produced longer scripts than their TD peers) and mark (to ascertain whether the SpLD candidates achieved higher marks than their TD peers). This was based on the premise that additional time results in longer answers and longer answers (as they include a greater wealth of detail and more fully address the question) result in higher marks (Zuriff, 2000).

Duncan & Purcell (2017) found that students with SpLD who were granted extra time in their exams did not produce a higher word count or achieve higher marks than their TD peers who took the same exam under standard conditions. In addition, the study identified no significant correlation between word count and mark across the participants as a whole, with higher word count failing to correlate with higher mark and low word count failing to correlate with low marks. These outcomes are consistent with the research of Wadley et al., (2013) and Gregg & Nelson (2010) who found that longer answers did not improve the test scores of either SpLD or TD participants.

Although the research outcomes from this study demonstrated that there were no statistically significant differences in exam performance between the SpLD participant group as a whole and their TD peers, some differences were noted between the SpLD sub-groups and their TD comparators (Duncan & Purcell, 2017). For example, when analysing differences in exam performance between the SpLD and TD groups, the SpLD group who had 25% extra time alone achieved lower marks than their TD peers who took the same exam under standard conditions. However, no statistical differences in exam performance were apparent between the SpLD group who used a word processor in addition to extra time and the TD group, either in terms of word count or mark. Therefore, this evidence suggests that the use of a word processor in addition to extra time levels the playing field for students with SpLD in a way that extra time alone fails to do. This outcome reflects the research of Elliott & Marquart (2004) and Fuchs & Fuchs (2001) which concluded that *'other factors improve students' test scores more significantly than additional time'* (Elliott & Marquart, 2004, p365), and *'that students with disabilities require more than the single accommodation of extended time'* (Fuchs & Fuchs, 2001, p81).

9.2 Variability in the amount of extra time granted

A further difficulty in drawing a unified conclusion from the research in this literature review and applying it to the question of extra time in school or university exams relates to a lack of consistency in the amount of extra time applied across the studies. Reflecting differences in local practices, there is variability across the studies cited above in terms of the amount of extra time granted to the participants and this may contribute to the contradictory findings. For example, Runyan (1991) granted unlimited extra time, Elliott & Marquart (2004) allowed 100% additional time, Lewandowski, Cohen & Lovett (2013) awarded 50% and 100% extra time, while in the study by Duncan & Purcell (2017) the participants were granted 25%. This lack of uniformity may undermine the integration of conclusions about the effect of extra time across studies. That is to say, where it is found that extra time improves the performance of all students, rather than promoting equity across all candidates, the degree to which the test scores were inflated (and therefore the validity of these findings) may be influenced by the amount of extra time that was implemented. The differences in amount of extra time awarded in the studies (coupled with differences in inherent 'speededness' of the tests undertaken) render it difficult to draw an overall conclusion from the findings as a whole.

10. Conclusion

The studies included in this literature review all aimed to investigate the effects of changing the standardised administration of the exam in terms of the time allocation. The objective was to determine whether or not the granting of extra time simply compensates for the student's disability and removes the barrier that they would otherwise experience in being able to perform to their potential, or if it benefits all candidates and so undermines the validity of the exam by over-inflating scores.

It is clear from this review that there exists an insufficient body of non-contradictory, reliable research studies investigating impact of granting extra time in exams to students with SpLD. There is no consensus of opinion as to whether or not extra time benefits all students, or just students with SpLD. This reflects the findings of The National Centre on Educational Outcomes (NCEO), which similarly concluded that '*the findings about the effect of extended time on assessment scores varied widely, yielding no consensus*' (NCEO, 2015, p18).

However, despite the inconsistencies in findings between studies a pattern does emerge across all the studies as a whole, which suggests that the differences in outcomes may be linked to the inherent 'speededness' of the tests used in the study. For example, those studies that used 'high speeded' tests found that all participants improved with extra time. This may be due to the study not allowing sufficient time for any candidate to complete the test under standard time conditions (so all participants improved with extra time due to the removal of any test ceiling). Those studies that were 'moderately speeded', such that the standard time was allocated so that some, but not all, TD candidates were expected to complete the exam without needing extra time resulted in a differential boost. That is to say, all candidates improved their score with the extra time, but those with SpLD showed greater gains. Those studies that determined standard time in such a way as to allow the vast majority of candidates to complete the exam in standard time showed that only candidates with SpLD increased their scores when granted extra time, suggesting that only those candidates who were not able to complete the exam in standard time benefitted from further time. Finally, those studies

that overestimated the amount of standard time needed to complete the exam found that no groups improved their scores when given more time. This may be because all candidates comfortably completed the test within the standard time (due to a low test ceiling), with the result that there were no further questions remaining unanswered from which to improve their score during the ‘extra time’ period. Figures 1 & 2 provide a diagrammatical depiction of how the differences in inherent ‘speededness’ of the tests used in the different research studies impact on the outcomes.

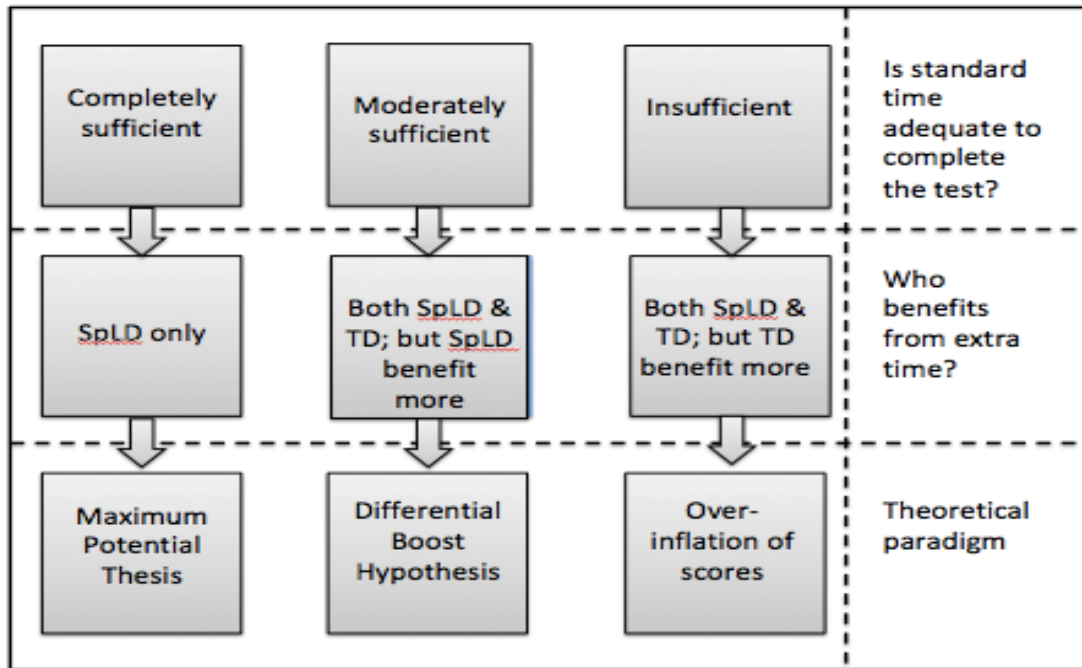


Figure 1: Relationship between the amount of standard time allocated in tests and the effect of extra time on participants’ scores

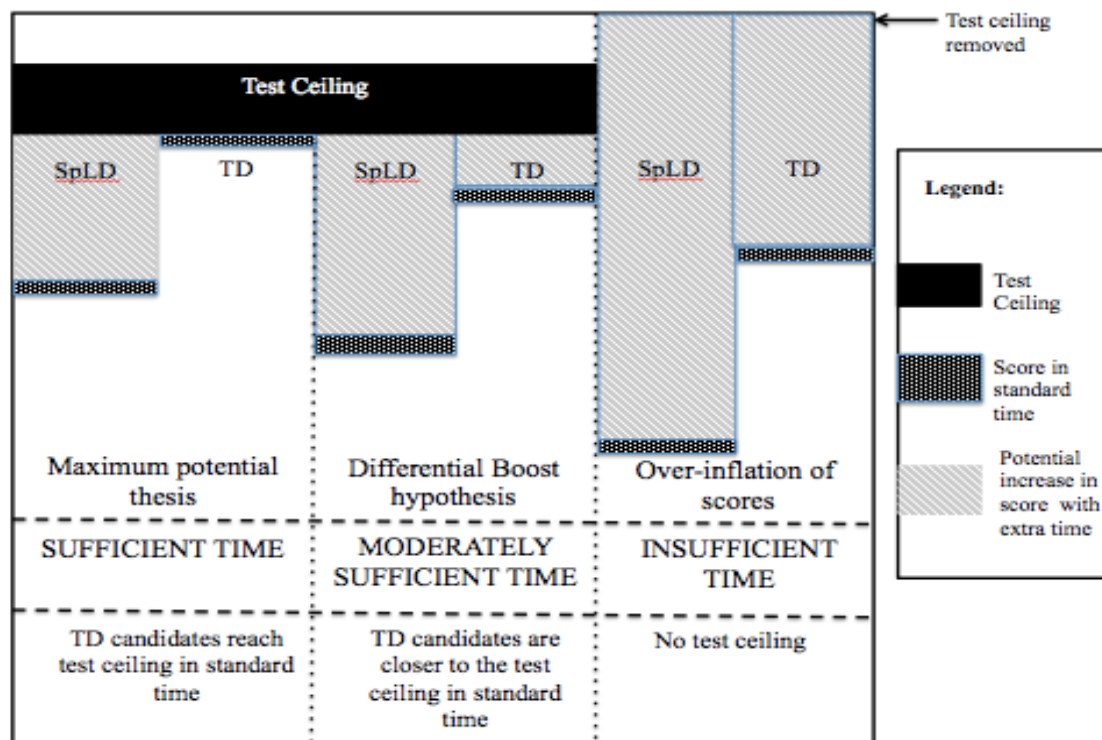


Figure 2: Relationship between inherent ‘speededness’ of the test and the potential increase in score that results from granting extra time.

This explanation is supported by the conclusion of the systematic review of literature conducted by Sireci, Scarpati & Li (2005), which comments that:

“Extended-time administrations provide a more precise estimate of students’ abilities than standard-time administrations, and differences in scores change are more related to the effect of speed on regular administrations, rather than providing learning disabled students with any advantage due to extended time.” (Sireci, Scarpati & Li, 2005, p469)

Despite the differences in outcomes across the body of literature, where the studies do converge is in the findings that TD students perform better on timed tests than students with SpLD (especially under standard conditions) and that the granting of extra time demonstrates a positive effect on test scores for students with SpLD. This suggests that extra time benefits students with SpLD and helps to normalise their performance to that of their TD peers. Nonetheless, if all students benefit from additional time this raises the question that the extra time is simply increasing the score of the candidate in receipt of extra time, rather than achieving a more accurate score that is a valid representation of the student’s subject knowledge and skills.

Assimilating the findings of the research in this field into a unified conclusion is problematic, given the variety of methodologies and research designs used. As Gregg & Nelson (2012) comment:

‘...the body of research examined is not adequate to address many of the questions pertaining to the comparability of scores with and without [exam] accommodation’ and ‘the literature is lacking in quantity of studies, restricted

in types of design methodologies and under-representative of the diversity of individuals demonstrating the disorder' (Gregg & Nelson, 2012, p142).

As a result, although there is general agreement in the field that additional time improves outcomes for students with SpLD, there is no unified view about the fairness of this adjustment (i.e. whether it is only students with SpLD who benefit or whether all students benefit similarly from additional time). This demonstrates a clear need for further studies in this area in order to enable valid conclusions to be drawn about the comparability of results for students with SpLD taking exams with additional time to TD peers who take those same exams under standard conditions.

References

- Bolt, S. E., Decker, D. M., Lloyd, M., & Morlock, L. (2011). Students' perceptions of accommodations in high school and college. *Career Development for Exceptional Individuals*, 34(3), 165-175. DOI: [10.1177/0885728811415098](https://doi.org/10.1177/0885728811415098)
- Brown, J.A., Fisco, V.V., & Hanna, G. (1993). *Nelson-denny reading test: Manual for scoring and interpretation, forms G & H*. Rolling Meadows, IL: Riverside Publishing
- Bryman, A. , Teevan, J., Bell, E. 2009. *Social Research Methods*. (2nd Ed.). Oxford University Press. Oxford.
- Cohen, A., Gregg, N., & Deng, M. (2005). The role of extended time and time content on a high-stakes mathematics test. *Learning Disabilities Research and Practice*. 20
- Duncan, H., & Purcell, C. (2017). Equity or Advantage? The effect of receiving exam access arrangements in University exams on Humanities students with SpLD. *Widening Participation and Lifelong Learning*. 19 (2)
- Elliott, S. N., Marquart, A.M. (2004). Extended time as a testing accommodation: Its effects and perceived consequences. *Exceptional Children*. 70 (3). pp349 – 367.
- Fuchs, L. S. & Fuchs, D. (2001). Helping Teachers Formulate Sound Testing Accommodation Decisions for Students with Learning Disabilities. *Learning Disabilities: Research & Practice*. 16 (3). pp174 – 181
- Frederickson, N., & Reason, R. (1995). Discrepancy Definitions of Specific Learning Difficulties. *Educational Psychology in Practice*. 10 (40). pp195-205
- Gregg, N., & Nelson, J.M. (2012). Meta-analysis on the Effectiveness of Extra Time as a Test Accommodation for Transitioning Adolescents with Learning Disabilities: More Questions than Answers. *Journal of Learning Disabilities*. 45 (2) pp128-38 doi: 10.1177/0022219409355484.
- Lesaux, N.K., Pearson, M.R., & Siegel, L.S. (2006). The effects of timed and untimed testing conditions on the reading comprehension of adults with reading disabilities. *Reading and Writing*. 19
- Lewandowski, L. J., Cohen, J., & Lovett, B. J. (2013). Effects of Extended Time Allotments on Reading Comprehension Performance of College Students With and Without Learning Disabilities. *Journal of Psychoeducational Assessment*, 31(3), 326-336. DOI: 10.1177/0734282912462693
- Lovett, B.J. (2011). Extended Time Accommodations: What does the Research say? *Communique: The newspaper of the National Association of School Psychologists (NASP)*. 39 (1)
- National Centre on Educational Outcomes (2016). *Assessment Accommodations use*

by Students Receiving Special Education Services. *NCEO APR Snapshot Briefs*.
11. National Centre on Educational Outcomes (NCEO). Retrieved from
<https://ici.umn.edu/index.php?products/view/894> Accessed 30.5.17.

Phillips, S.E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurements in Education*. 7 (2).

Reid, G. (2009). *Dyslexia: A Practitioner's Handbook*. Wiley-Blackwell. Chichester.

Runyan, M. K. (1991). The Effect of Extra Time on Reading Comprehension Scores for University Students With and Without Learning Disabilities. *Journal of Learning Disabilities*. 24 (2)

Sireci, S. G. & Parker, P. (2005). Validity on trial: Psychometric and legal conceptualizations of validity. *Educational Measurement: Issues and Practice*. 25. pp27–34.

Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*. 75 (4). pp 457-490

Zuriff, G.E. (2000). Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Applied Measurement in Education*. 13