BABCP
www.babcp.com

MAIN

An evaluation of the psychometric properties of the adapted PHQ-9 and GAD-7 outcome measures for use with adults with intellectual disability

Hannah Jenkins[1] , Kate Theodore[2] , Myra Cooper[1], Jennifer Breen[2] and Olivia Hewitt[3,4]

[1]The Oxford Institute of Clinical Psychology Training and Research, Isis Education Centre, Warneford Hospital, Oxford, UK, [2]Department of Psychology, Royal Holloway University of London, Egham, UK, [3]Berkshire Healthcare NHS Foundation Trust, Psychology Service, Bracknell, UK and [4]University of Birmingham, Health Services Management Centre, Birmingham, UK

Corresponding author: Olivia Hewitt; Email: OMH461@student.bham.ac.uk

Abstract

Background: People with intellectual disability often face barriers accessing mainstream psychological services due to a lack of reasonable adjustments, including the absence of adapted versions of routine outcome measures. Adapted versions of the Patient Health Questionnaire-9 (PHQ-9) and the Generalised Anxiety Disorder-7 (GAD-7) have been created for adults with ID.
Aims: This study aims to evaluate the psychometric properties of the adapted PHQ-9 and GAD-7.
Method: The adapted PHQ-9 and GAD-7 and the Glasgow Depression and Anxiety Scales (GDS-ID, GAS-ID) were administered to 47 adults ($n = 21$ clinical group; $n = 26$ community group) with ID. Cross-sectional design and between-group analyses tested for discriminant validity. Concurrent and divergent validity was tested using correlational designs. Reliability was investigated by internal consistency and test–retest analysis.
Results: The clinical group scored significantly higher on the adapted PHQ-9 ($t_{45} = -2.28$, $p = .03$, 95% CI [–7.09, –.45]) and GAD-7 ($t_{45} = -3.52$, $p = .001$, 95% CI [–7.44, –2.02]) than the community group, evidencing discriminant validity. The adapted PHQ-9 correlated with the GDS-ID ($r_{47} = .86$, $p < .001$) and the adapted GAD-7 correlated with the GAS-ID ($r_{46} = .77$, $p < .001$). The adapted PHQ-9 (Cronbach's $\alpha = .84$, ICC = .91) and GAD-7 (Cronbach's $\alpha = .86$, ICC = .77) had good internal consistency and test–retest reliability.
Conclusions: Preliminary research suggests the adapted PHQ-9 and GAD-7 are valid and reliable measures. They could provide a reasonable adjustment for the minimum dataset used in NHS Talking Therapies and can be easily administered in routine clinical practice. Further work to establish additional psychometric properties is now required.

Keywords: GAD-7; intellectual disability; outcome measures; PHQ-9; psychometric properties; reasonable adjustments

Introduction

Background

Intellectual disability and mental health

An estimated 1.5 million people, or 2.16% of adults in the UK have an intellectual disability (MENCAP, n.d.). Individuals with intellectual disability have significant impairments of

intellectual functioning and adaptive behaviour (everyday living skills and social functioning), with both impairments arising before adulthood (British Psychological Society, 2015). They have difficulties with communication and require support in activities of daily living.

Adults with intellectual disability are at increased risk of mental health difficulties due to multiple psychosocial pre-disposing factors. Despite interventions that promote social inclusion, people with intellectual disability feel excluded and isolated (Merrells *et al.*, 2018). They have limited employment opportunities and are economically disadvantaged, which is associated with poor mental health (Dean *et al.*, 2018) and reduced quality of life (Simões and Santos, 2016). People with intellectual disability have more restricted and impoverished social relationships (Lippold and Burns, 2009), which, alongside social deprivation and experiences of failure, increases mental ill health and reduces quality of life. Additionally, people with intellectual disability are significantly more likely to experience adverse life events, leading to symptoms of trauma (Wigham *et al.*, 2014) and are more vulnerable to abuse and trauma than the general population (McDonnell *et al.*, 2019).

The label 'intellectual disability' is highly stigmatising and can result in discrimination (MENCAP, 2018), poor self-esteem and increased vulnerability (Jahoda and Markova, 2004; Paterson *et al.*, 2012). The awareness of stigmatisation can threaten identity and sense of well-being (Jahoda and Markova, 2004). Cognitive behaviour theory proposes that negative self-evaluations contribute to the development of emotional problems (Beck, 2002). Internal factors also contribute to mental health difficulties in people with intellectual disability, including higher levels of anxiety (in particular intolerance of uncertainty), difficulties recognising and regulating emotions, sensory sensitivities, and poor executive functioning (e.g. Kiani and Miller, 2010; Sáez-Suanes *et al.*, 2020).

Given the multiple pre-disposing factors and vulnerabilities, it is unsurprising that higher prevalence of mental health problems are seen in adults with intellectual disability (23.4% compared with 5.3% of adults in the general population) (Hughes-McCormack *et al.*, 2017). However, there are multiple challenges to identifying mental health problems within intellectual disability populations. Communication difficulties can make it harder for individuals to explain their experiences and for professionals to understand their problems – potentially preventing individuals from seeking help and accessing services. Detailed assessments by specialised professionals can be required, as mental health difficulties can present differently in people with intellectual disability. Co-morbidities and physical health problems and/or sensory and cognitive difficulties can mask mental health problems (Moss *et al.*, 2000; NICE, 2016). Additionally, individuals with moderate to severe intellectual disability may present with challenging behaviours that can be attributed to their intellectual disability as opposed to a possible underlying mental health difficulty (Moss *et al.*, 2000). Furthermore, there are limited clinical measures available to detect mental health difficulties in intellectual disability populations (Vlissides *et al.*, 2016). Such measures may lack accessibility, and, if not developed with sufficient input from people with intellectual disability, may lack acceptability. The National Institute of Health and Care Excellence (NICE, 2016) reports that no suitable tools exist for routine use in primary care settings to identify common mental health difficulties in intellectual disability populations.

Psychological services, reasonable adjustments, and outcome measures

The Mental Health National Service Framework (Department of Health, 1999) adopts an inclusive approach and key policies (e.g. Valuing People; Department of Health, 2001 and The Greenlight Toolkit; NDTi, 2016) state that individuals with an intellectual disability should be able to access mainstream services. Psychological services (e.g. NHS Talking Therapies, formerly known as Improving Access to Psychological Therapies [IAPT]) have a legal responsibility to ensure people with protected characteristics, as outlined in the Equality Act, have equal access to care (HM Government, 2010). Services are required to provide 'reasonable adjustments' to ensure people

with protected characteristics are not discriminated against. However, people with a disability (a protected characteristic) are under-represented in mainstream psychological services (Department of Health, 2012) and face barriers to accessing psychological support. Whilst specialist services provided through Community Learning Disability Teams (CLDTs) are essential in providing mental health care for people with intellectual disability, they should use the same services, resources, and facilities as the rest of the general population (Department of Health, 2001). Individuals with intellectual disability who can access mainstream services, with some reasonable adjustments, should be supported to do so.

Outcome measures are routinely used in many psychological services, typically to assess and monitor progress between treatment sessions, and to evidence service delivery and effectiveness. The Patient Health Questionnaire (PHQ-9; Kroenke *et al.*, 2001), a self-report measure for depression, and the Generalised Anxiety Disorder scale (GAD-7; Spitzer *et al.*, 2006) are standard outcome measures, which form part of the minimum dataset, used in NHS Talking Therapies (the main UK primary care psychological service).

While these measures are well validated for use with the general population, they may not be suitable for people with intellectual disability (Chinn *et al.*, 2014) particularly those with difficulties in reading, writing, and comprehension. People with and without intellectual disabilities accessing Talking Therapies have broadly similar patterns regarding waiting times and proportion of people finishing treatment, and clinical outcomes are poorer for people with intellectual disabilities (Dagnan *et al.*, 2022). The IAPT Learning Disabilities Good Practice Guide (Dagnan *et al.*, 2015) acknowledges the difficulties of using the GAD-7 and PHQ-9 with patients with intellectual disabilities, stating 'It is highly likely that people with difficulties in completing the measures who are not given support to do so will drop out of therapy'. Various adaptations are detailed that may be necessary to make these measures accessible including reading the questionnaire aloud to the person, changing the wording on the questionnaires, and breaking down questions with multiple components to deliver one element at a time. That the psychometric properties of the GAD-7 and PHQ-9 have not been established in the intellectual disability population is a limitation by researchers investigating the ability of Talking Therapy services to effectively treat people with intellectual disabilities (e.g. Whittaker, 2023). Therefore, the absence of adapted versions of routine outcome measures (e.g. the PHQ-9 and GAD-7) is a barrier for people with intellectual disability accessing Talking Therapies services and evidence-based interventions.

Current outcome measures available for people with intellectual disability

Measures to identify and monitor changes in mental health difficulties have been developed for individuals with intellectual disability; however, there is no 'gold standard' consistently used across services (Patel *et al.*, 2023). Having consistent clinical outcome measures are important to understand the efficacy of treatments provided by all services – at an individual, service, and national level. Having valid and consistent measures will allow services to better track recovery rates for patients with an intellectual disability. Existing measures have been developed using small sample sizes and therefore, are not well validated (British Psychological Society, 2015; McGurk and Skelly, 2014). The lack of valid and reliable measures makes evaluating the effectiveness of psychological interventions for people with intellectual disability challenging.

Current measures used in specialist intellectual disability services include the Psychological Therapies Outcome Scale (PTOS-ID; Vlissides *et al.*, 2017), the Clinical Outcome Routine Evaluation – Learning Disabilities (CORE-LD; Barrowcliff *et al.*, 2018), the Glasgow Depression Scale (GDS-ID; Cuthill *et al.*, 2003) and the Glasgow Anxiety Scale (GAS-ID; Mindham and Espie, 2003). The PTOS-ID aims to measure psychological distress and psychological well-being. It has good reliability and good levels of construct and concurrent validity; however, further amendments and psychometric testing are required (Vlissides *et al.*, 2017). The CORE-LD has

a 30- or 14-item version and is a useful broad ranging measure of psychopathology, with good levels of internal consistency and convergent validity (Barrowcliff *et al.*, 2018). The GDS-ID has specifically been created as a self-report measure of depressive symptoms for people with intellectual disability. It has good internal consistency, test–retest reliability, and acceptable face and content validity (Cuthill *et al.*, 2003). The GAS-ID (Mindham and Espie, 2003) has been developed to assess for anxiety problems in people with intellectual disability and is psychometrically robust. Whilst the Glasgow scales are the most established measures to assess depression and anxiety in intellectual disability populations, they are not fully accessible, are time-consuming to complete and are not currently used in mainstream psychological services.

NICE (2016) and the Department of Health (2008) recommend adapting the minimum dataset used in Talking Therapies services, thus ensuring reasonable adjustments allow adults with intellectual disability equitable access to mainstream services providing evidence-based psychological therapies. Valid and reliable adapted outcome measures for individuals with intellectual disability will help with early intervention and likely provide better outcomes. Adapted versions of the PHQ-9 and GAD-7 have been designed specifically for individuals with intellectual disability, using a concurrent verbal probing cognitive interviewing methodology adapted for people with intellectual disability (Breen, 2017). This involved an iterative process of examining adapted versions of the measures, and seeking detailed feedback from people with intellectual disability, paid support staff and clinicians, through individual interviews, focus groups and written feedback. These measures comply with easy-read guidelines, using images and pictures in addition to text, and fulfil the recommendations proposed in the Learning Disabilities Positive Practice IAPT Guidelines (Dagnan *et al.*, 2015). Initial psychometric investigations were encouraging but limited in scope, and further psychometric investigation is warranted before services adopt these measures (Breen, 2017).

Aims

This study aimed to investigate the psychometric properties of the adapted PHQ-9 and GAD-7. Specifically, we aimed to evaluate the measures' internal consistency and test–retest reliability. We aimed to test for concurrent validity through comparisons with the Glasgow depression and anxiety measures and to test for divergent validity. We aimed to test for discriminant validity by comparing scores on the measures between a clinical and non-clinical (community) sample, predicting that the clinical group would score significantly higher on the measures than the community group.

Method

Participants

Participants were all adults (aged 18 years and above) with a mild to moderate intellectual disability with capacity to consent to research. The clinical group were patients of Community Teams for People with Learning Disabilities (CTPLDs) accessing support from services for a mental health difficulty and were recruited from five NHS Trusts across England. The community group included participants who were not currently accessing any mental health services, and were recruited by attending various virtual advocacy groups, day centres, colleges, and supported living services, and through emailing colleagues and sharing social media posts and adverts. Due to COVID-19 restrictions, all recruitment took place remotely, therefore access to video conferencing platforms (with support if required) became an inclusion criterion. Approximately half of all participants required additional support from family members or carers to access the study and the technology required. Anyone unable to speak English or communicate verbally, or who had a diagnosis of dementia, was excluded.

Table 1. Demographic characteristics of participants

	Group		
	Clinical	Community	Total (%)
Gender			
Male	9	16	25 (53.2)
Female	12	9	21 (44.7)
Other	0	1	1 (2.1)
Ethnicity			
White	18	22	40 (85.1)
Black	1	2	3 (6.4)
Asian	1	1	2 (4.3)
Other	1	1	2 (4.2)
Age			
20–29	10	8	18 (38.3)
30–39	2	12	14 (29.7)
40–49	4	4	8 (17.1)
50–60	5	2	7 (14.9)
Total	21	26	47
Additional co-morbidities			
Autism spectrum disorder	11	5	
ADHD	1	3	
Down syndrome		1	
Dandy-Waller syndrome		1	
Williams syndrome		1	
Cerebral palsy		1	
Epilepsy	2	1	
Heart problems	2	2	

Forty-eight adults completed questionnaires between October 2020 and June 2021. One participant was excluded (see below for details) making a total of 47 participants. They had a mean age of 35.15 years ($SD = 10.52$) with a range of 20–60 years. There were 25 males, 21 females and one participant identified as 'other'. There were 21 participants in the clinical group and 26 in the community group. The groups did not differ significantly in age, ethnicity, or gender distribution. Table 1 presents demographic characteristics, including co-morbid health conditions self-reported by participants.

Design

A quantitative evaluation of the psychometric properties of the adapted PHQ-9 and GAD-7 was conducted using a cross-sectional design and between-group analyses to test for discriminant validity. Concurrent and divergent validity was tested using a correlational design. Opportunity sampling was used to recruit participants to either the clinical or community group. Questionnaires were completed at a single time point for participants in the clinical group. In the community group, the adapted PHQ-9 and GAD-7 were repeated at a second time point to assess test–retest reliability.

Measures

Glasgow Depression Scale (GDS-ID; Cuthill et al., 2003)

The GDS-ID is a self-report measure of depression specifically developed for adults with mild to moderate ID. It takes 15 minutes to administer and consists of 20 items that are rated on a 3-point Likert scale: 'never', 'sometimes', 'always'. The suggested clinical cut-off is a score of 13 or more. Five items are reverse scored, before adding up individual scores to give a total score (out of 40). It is reported to be a sensitive measure, with good internal consistency (Cronbach's alpha .9) and good test–retest reliability ($r = .97$) (Cuthill *et al.*, 2003).

Glasgow Anxiety Scale (GAS-ID; Mindham and Espie, 2003)

This is a self-report measure of anxiety specifically developed for adult intellectual disability populations and takes between 5 and 10 minutes to complete. Items are rated on the same Likert scale as the GDS-ID, with a clinical cut-off of 13. One item is reverse scored, and individual scores are added together to give a total score out of 54. It consists of 27 items divided into questions about 'worries', 'specific fears' and 'physiological symptoms'. The GAS-ID has good internal consistency (Cronbach's alpha .96) and good test–retest reliability ($r = .95$).

Adapted PHQ-9 (Breen, 2017)

The adapted PHQ-9 is a 9-item self-report measure of depression for adults with intellectual disability. It has a 4-point Likert scale, with the following responses: 'no days', 'some days', 'a lot of days', 'nearly every day'. The measure has been adapted from the PHQ-9 (Kroenke *et al.*, 2001) by a process of cognitive interviewing to ensure the measure is more accessible for adults with intellectual disability. The measure has been adapted by including pictures with easy-to-read English, prompts for the facilitator and rewording of the scale. It has been estimated to take between 5 and 10 minutes to complete (Breen, 2017). Individual item scores are summed to give an overall score (out of 27), with higher scores indicating more severe depression symptoms.

Adapted GAD-7 (Breen, 2017)

This is a 7-item self-report measure of anxiety for adults with intellectual disability. It has the same 4-point Likert scale as the adapted PHQ-9 and has been adapted from the GAD-7 (Spitzer *et al.*, 2006) by a process of cognitive interviewing with adults with intellectual disability and has a similar format to the adapted PHQ-9. It takes about 5 minutes to complete (Breen, 2017). Individual item scores are summed to give an overall score (out of 21) with higher scores indicating higher anxiety levels.

Procedure

Easy-read information sheets were designed, with support from service users, to explain the research. Participants gave verbal informed consent (due to the study being conducted via video conferencing) and received electronic copies of their consent form. All participants in the community sample completed questionnaires via video conferencing with H.J. Participants in the clinical group completed questionnaires with a suitably qualified clinician in their service ($n = 16$) or with H.J. ($n = 5$). Questionnaires were all administered in the same order (demographic questionnaire, adapted PHQ-9, adapted GAD-7, GDS-ID, GAS-ID) as no order effects were expected. Participants were informed it was not a test and there were no right or wrong answers. Participants were informed they would be asked questions relating to their feelings in the past week, and the rating scale was explained to them. Questions were read out verbatim and the 'share screen' function was used to show each question visually. Participants in the community sample attended a second meeting around a week later to complete the adapted PHQ-9 and GAD-7 again. Verbal informed consent was obtained again at the second visit. Participants were debriefed and were offered a certificate of participation, thanking them for their contribution.

Data analysis

The data were analysed using SPSS 27.0. Parametric test assumptions were checked prior to conducting analysis. Cronbach's alpha was used to measure the internal consistency of each outcome measure. Test–retest reliability was assessed through intraclass correlation coefficients (ICC) as it provides the degree of correlation and agreement between measurements (Koo and Li, 2016). Independent *t*-tests were computed to test for discriminant validity, to see if there were any

significant differences between clinical and community group scores on the measures. Concurrent validity was tested by correlating the adapted PHQ-9 with the GDS-ID and the adapted GAD-7 with the GAS-ID. Divergent validity was tested by correlating the questionnaires with age and gender. One participant from the clinical sample was excluded from analysis as they answered 'zero' on every item of each questionnaire. It was assumed the participant did not understand the measures as it was reported they were accessing weekly psychology sessions due to anxiety and panic attacks. There was one missing value on the GAS-ID (question 10) for one of the clinical group participants – this participant's total score on the GAS-ID was not included in the discriminant validity analysis and only total subscale scores were calculated for the 'Specific fears' and 'Physiological symptoms' and not the 'Worry' subscale for this participant. No other data were missing from the sample.

Results

Histograms, normal Q-Q plots and box plots were visually inspected for the entire sample and the questionnaire scores appeared to be normally distributed. When the data was split into the clinical and community groups, the adapted PHQ-9, GAD-7, and GDS-ID appeared slightly positively skewed; however, analysis of skewness and kurtosis indicated that the measures were within the normal limits (z-values were between -1.96 and $+1.96$), thus the data were suitable for parametric testing.

Reliability

Internal consistency
Cronbach's alpha for the adapted PHQ-9 ($n = 47$) was .84, suggesting good internal consistency (Kline, 2000). The range in internal consistency for the total scale, measured by alpha if item deleted was between .81 and .85. The item total correlations ranged from .30 to .74.

Cronbach's alpha was .86 for the adapted GAD-7 ($n = 47$), with the range in internal consistency, as measured by alpha if item deleted, being .82 to .85. The item total correlations ranged from .53 to .73. The measure thus showed good internal consistency (Kline, 2000).

Cronbach's alpha for the GDS-ID ($n = 47$) was .91 and for the GAS-ID ($n = 46$) was also .91. Both Glasgow measures showed excellent reliability (Kline, 2000).

Whilst comparing the Cronbach's alpha for the adapted PHQ-9 and GAD-7 with the original measures does not provide strong evidence for validity of the adapted measures, it allows for comparison in patterns of data. Cronbach's alpha for the adapted PHQ-9 ($n = 47$) was .84, compared with a Cronbach's alpha of 0.89 (Kroenke *et al.,* 2001) for the original measure. Cronbach's alpha was 0.86 for the adapted GAD-7 ($n = 47$), while the unadapted GAD-7 has a reported Cronbach's alpha of 0.92 excellent internal consistency (Spitzer *et al.,* 2006). We report a correlation of 0.77 between the adapted PHQ-9 and adapted GAD-7 (see Table 2), with correlations of unadapted PHQ-9 and GAD-7 reported as 0.73–0.87, $p < 0.005$ (Shah *et al.,* 2021).

Test–retest reliability
Data on 24 test–retest sets were available. The mean time taken between each administration of the adapted PHQ-9 and GAD-7 was 9 days ($SD = 2.82$; range 7–17 days). Two-way mixed ICCs with absolute agreement were calculated for test–retest reliability. An ICC estimate of .40 is poor, .40–.59 is fair, .60–.74 is good and .75–1.0 suggests excellent reliability (Cicchetti, 1994). An ICC estimate of 1 indicates perfect agreement and 0 indicates random agreement. An excellent degree of test–retest reliability was found for the adapted PHQ-9 (ICC = .91, 95% CI .80–.96, $p < .001$) and for the adapted GAD-7 (ICC = .77, 95% CI .55–.90, $p < .001$).

Table 2. Means, standard deviations, and correlations

Measure	M	SD	PHQ-9	GDS-ID	GAD-7	GAS-ID	GAS-ID Worries	GAS-ID Specific fears	GAD-ID Physiological symptoms
PHQ-9	8.30	5.87	1						
GDS-ID	12.96	8.51	.86**	1					
GAD-7	7.38	5.12	.77**	.81**	1				
GAS-ID	19.30	10.83	.80**	.82**	.77**	1			
GAS-ID Worries	8.54	4.90	.74**	.80**	.75**	.89**	1		
GAS-ID Specific fears	5.30	3.86	.51**	.46**	.52**	.82**	.57**	1	
GAS-ID Physiological symptoms	5.51	3.82	.78**	.81**	.67**	.85**	.65**	.56**	1

M and *SD* are used to represent mean and standard deviation, respectively. **Correlation is significant at the 0.01 level (2-tailed).

Table 3. Mean total scores and standard deviations

	Group		
Measure	Clinical	Community	Total groups
Adapted PHQ-9	10.38 (5.80)	6.62 (5.47)	8.30 (5.87)
Adapted GAD-7	10.00 (4.73)	5.27 (4.47)	7.38 (5.12)
GDS-ID	16.14 (7.38)	10.38 (8.61)	12.96 (8.51)
GAS-ID	24.50 (11.02)	15.31 (8.97)	19.30 (10.83)

Validity

Discriminant validity

The ability of the adapted PHQ-9 and GAD-7 measures to discriminate between the clinical group and the community group was investigated by conducting independent *t*-tests. Table 3 reports group means and standard deviations for each measure.

On average, the clinical group scored higher on the adapted PHQ-9 ($M = 10.38$, $SD = 5.80$, $n = 21$) than the community group ($M = 6.62$, $SD = 5.47$, $n = 26$). There was a significant difference in depression scores between groups, $t_{45} = -2.28$, $p = .03$, 95% CI [–7.09, –.45]. According to Cohen's (1988) conventions the effect size was medium to large (Cohen's $d = .65$).

The clinical group also scored higher on the adapted GAD-7 ($M = 10.00$, $SD = 4.73$, $n = 21$) than the community group ($M = 5.27$, $SD = 4.47$, $n = 26$). There was a significant difference in anxiety scores between groups, $t_{45} = -3.52$, $p = .001$, 95% CI [–7.44, –2.02]. The effect size was large (Cohen's $d = 1.03$).

We also tested the Glasgow measure's ability to discriminate between the clinical and community groups. The clinical group scored higher on the GDS-ID ($M = 16.14$, $SD = 7.38$, $n = 21$) than the community group ($M = 10.38$, $SD = 8.61$, $n = 26$) and there was a significant difference between the two groups, $t_{45} = -2.43$, $p = .019$, 95% CI [–10.54, –.98], $d = .72$.

The clinical group also scored higher on the adapted GAS-ID ($M = 24.50$, $SD = 11.02$, $n = 20$) than the community group ($M = 15.31$, $SD = 8.97$, $n = 26$). There was a significant difference in anxiety scores between groups, $t_{44} = -3.12$, $p = .003$, 95% CI [–15.13, –3.25], $d = .91$.

Concurrent validity

The clinical and community groups were combined ($n = 47$) for the correlational analysis. Concurrent validity, whereby a new measure is administered at the same time as a pre-existing
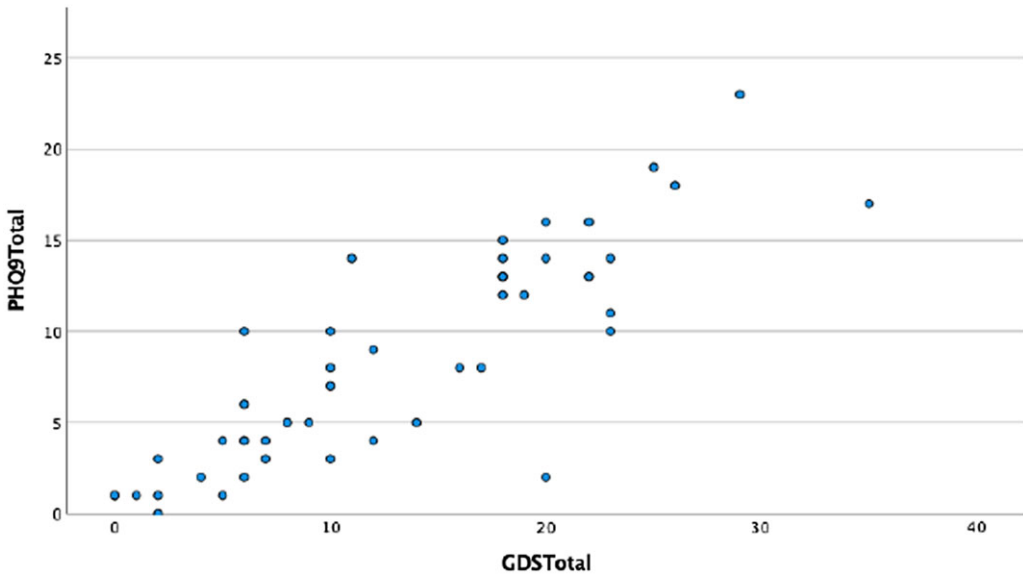
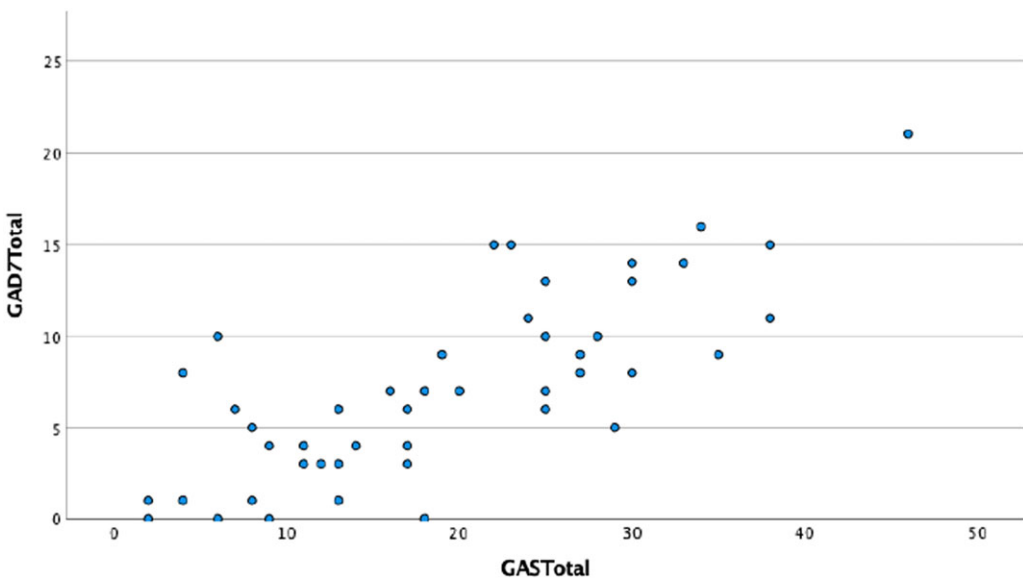**Figure 1.** Scatterplot of PHQ-9 Total by GDS Total.

Figure 2. Scatterplot of GAD-7 Total by GAS Total.

measure and the two are correlated, was assessed using scatter plots (see Figs 1 and 2) and Pearson's product moment correlation. The adapted PHQ-9 total score was correlated with the GDS-ID total score. A significant positive relationship was found, $r_{47} = .86$, $p<.001$.

The adapted GAD-7 total score was positively correlated with the GAS-ID total score, $r_{46} = .77$, $p<.001$. The adapted GAD-7 total score was also positively correlated with the GAS-ID 'Worry' subscale ($r_{46} = .75$, $p<.001$), the 'Specific fears' subscale ($r_{47} = .52$, $p<.001$), and 'Physiological symptoms' subscale ($r_{47} = .69$, $p<.001$).

Divergent validity

To test that constructs that should have no relationship do, in fact, not have any relationship, we correlated total scores for the adapted PHQ-9 and GAD-7 with age. As expected, there were no significant correlations, providing evidence for divergent validity. The adapted PHQ-9 total score was not correlated with age, $r_{47} = .03$, $p = .85$. The adapted GAD-7 total score was not correlated with age, $r_{47} = -.01$, $p = .97$. There were no significant correlations between age, $r_{47} = -.06$, $p = .68$, with the GDS-ID or with the GAS-ID., $r_{476} = .07$ $p = .63$

We conducted independent *t*-tests between gender and total scores on the adapted PHQ-9 and GAD-7 to further test for divergent validity. Females scored slightly higher on the adapted PHQ-9 ($M = 8.71$, $SD = 6.64$, $n = 21$) than males ($M = 7.96$, $SD = 5.39$, $n = 25$) but there was no significant difference in total scores on the adapted PHQ-9 between males and females, $t_{44} = .43$, $p = .67$, 95% CI [–2.82, 4.33].

Females also scored higher on the adapted GAD-7 ($M = 8.90$, $SD = 5.96$, $n = 21$) than males ($M = 6.12$, $SD = 4.13$, $n = 25$) but there was no significant difference in total scores on the adapted GAD-7 between males and females, $t_{34.68} = 1.81$, $p = .08$, 95% CI [–.34, 5.91].

Discussion

The aim of this study was to further evaluate the psychometric properties of the adapted PHQ-9 and GAD-7 for use with adults with intellectual disability. The results suggest that both measures appear reliable and valid, and are thus clinically useful. Previous initial investigations by Breen (2017) reported the validity of the adapted PHQ-9 and GAD-7 as inadequate, as only one validity test was computed. To account for this previous limitation, we aimed to compute three validity tests, namely discriminant validity, concurrent validity, and divergent validity. Regarding reliability, Breen (2017) reported internal consistency, but no other reliability tests were conducted. We computed test–retest reliability in addition to testing for internal consistency to further evaluate the measure's reliability. Despite challenges with recruitment due to COVID-19 we were able to recruit a respectable sample size and compute parametric statistical analysis. Our sample size was comparable to the psychometric investigations for the Glasgow measures ($n = 38$ for the GDS-ID, $n = 54$ for the GAS-ID) and larger than Breen's initial testing ($n = 32$).

In relation to reliability, both the adapted PHQ-9 and GAD-7 measures were rated as having good internal consistency and excellent test–retest reliability based on conventional criteria proposed by Kline (2000) and Cicchetti (1994). However, Koo and Li (2016) suggest confidence intervals should be used and not the ICC estimate itself when evaluating the level of test–retest reliability. Therefore, the adapted PHQ-9 is considered to have a good to excellent level of test–retest reliability, whereas the adapted GAD-7 has a moderate to excellent level of test–retest reliability. Cronbach's alpha for the adapted PHQ-9 and GAD-7 were compared with unadapted measures and found to have broadly similar levels. Comparison of correlations between the adapted and unadapted measures were also in line with each other. Overall, the findings suggest stability in measurement.

The adapted PHQ-9 and GAD-7 measures are deemed to be valid. As hypothesised, the measures could discriminate between the clinical and community sample in terms of the total score on the depression and anxiety measures. Participants in the clinical sample scored significantly higher on the adapted PHQ-9 and GAD-7 than participants in the community sample who were not currently accessing mental health services. However, whilst the adapted PHQ-9 and GAD-7 discriminated between community and clinical samples, we acknowledge that differences in mean scores between clinical and community groups are smaller than in the original papers detailing divergent validity for these two measures (Mindham and Espie, 2003; Cuthill *et al.*, 2003). The environmental context of the COVID-19 pandemic may have influenced baseline levels of anxiety and low mood within the population at the time of conducting this research. Additional work is required to establish possible cut-offs for clinical caseness for the adapted measures.

Divergent validity was evidenced by the measures not being significantly correlated with age or gender. Concurrent validity was evidenced by the adapted PHQ-9, and GAD-7 measures being significantly correlated with the more established Glasgow measures. As expected, the adapted GAD-7 was more strongly correlated with the GAS-ID 'worries' subscale compared with the 'specific fears' and 'physiological symptoms' subscales. Hinkle *et al.* (2003) provide interpretation guidance for correlation coefficients, suggesting .90 to 1.00 is very highly correlated, .70–.90 is high, .50–.70 moderate and .30–.50 is low. Using these guidelines, both the adapted PHQ-9 and GAD-7 were highly correlated with the Glasgow measures. Anastasi and Urbina (1997) comment that 'correlations should be moderately high but not too high. If the new test correlates too highly with an already available test, without such added advantages as brevity or ease of administration, then the new test represents needless duplication' (p. 127). However, the adapted PHQ-9 and GAD-7 are much quicker to administer than the Glasgow measures and participants informally commented they found them much easier to complete. They adhere to many of the recommendations identified by a recent systematic review into best practice when adapting self-report measures for people with intellectual disability (e.g. piloting draft versions, careful consideration of visual analogue scale, and providing specific guidance for administrators around difficult items; Kooijmans *et al.*, 2022). Thus, it appears the adapted measures have added advantages and could be an alternative to the Glasgow measures, particularly in busy clinical settings.

Areas for future research

Due to the scope of this study, we only reported on reliability and validity. Fitzpatrick *et al.* (1998) advise investigating the feasibility, acceptability, and responsiveness of measures in addition to reliability and validity when assessing the psychometric quality of outcome measures. From anecdotal evidence collected by H.J., the measures were quick and easy to administer. Participants understood the rating scale and questions, it was rare that the supplementary material to provide prompts for the questions or to explain a question was required. From this, and previous reports from Breen (2017), it suggests the measures are acceptable and feasible.

However, future research is required to replicate our findings and further explore the psychometric properties of these measures. This will need to evaluate the sensitivity of the measures to detecting change over time (responsiveness) and the validity of the possible clinical cut-offs used in calculating recovery. As reliable improvement and reliable deterioration are outcomes in Talking Therapies, the clinical meaningful change scores need to be established. It would be helpful for future research to consider the similarities and differences between the responses of people with intellectual disabilities on the adapted and unadapted PHQ-9 and GAD-7 measures, as part of further investigations into the psychometric properties of these measures. Additionally, due to many services offering digital/online assessments and interventions following changes in services resulting from the COVID-19 pandemic, it could be valuable to produce electronic versions of the measures. Again anecdotally, our study suggests it is feasible to complete these measures remotely via video call using the 'share screen' function. It was easy to enlarge the text and show one question at a time on the screen. However, about half of participants required support to access measures electronically, reflecting the conclusions of Kooijmans *et al.* (2022) that whilst computerised testing offers great advantages, there is a lack of research into the opportunities and pitfalls for people with intellectual disability.

Implications for psychological therapies services

Additionally, the adapted PHQ-9 and GAD-7 measures provide a reasonable adjustment to the minimum dataset that must be routinely collected in NHS Talking Therapies services. Having psychometrically robust routine outcome measures that are compatible with the minimum dataset

and acceptable and accessible for people with intellectual disability will help reduce discrimination and be one less barrier to receiving equal access to mainstream psychological services. It would be helpful for similar work to be undertaken to adapt other measures within the minimum dataset such as the Work and Social Adjustment Scale. Additionally, other anxiety disorder-specific measures could also be adapted to improve accessibility for people with intellectual disabilities. However, the introduction of adapted measures has implications for the comparisons that can be made between those patients using adapted and non-adapted measures.

The adapted GAD-7 and PHQ-9 described in this paper formalise the recommendations laid out in NICE (2016) and the Learning Disabilities Positive Practice IAPT Guidelines (Dagnan *et al.*, 2015). Implementing the adapted measures would ensure clinicians across NHS Talking Therapies services in the UK were administering measures in a consistent way as opposed to each service devising individual 'reasonable adjustments'. Furthermore, without the consistent use of outcome measures suitable for people with intellectual disability, improved tracking of intervention efficacy for intellectual disability populations will not occur. This poses challenges to improving the evidence base by conducting randomised controlled trials (RCTs) within this population, which is warranted considering the high prevalence of mental health difficulties in this population.

Limitations

A limitation of this study was the small sample size and thus some caution is required when interpreting findings as there is the risk results are biased, consequently impacting on validity. Additionally, as it was a convenience sample, it may not be representative of the population, thus issues with generalisability arise. However, the sample was larger than the sample sizes used to evaluate the GDS-ID and GAS-ID and other self-report measures used in ID studies (e.g. mini-MANS-LD; Raczka *et al.*, 2018).

Conclusions

The present study has evaluated the psychometric properties of the adapted PHQ-9 and GAD-7 for use with adults with intellectual disability. The measures have shown to be reliable and valid. They provide a reasonable adjustment for the minimum dataset used in NHS Talking Therapies services and can be easily administered. However, further evaluation is required to establish additional properties of the measures before they can be adapted into routine clinical practice.

Ethical standards. The research was conducted in accordance with the Ethical Principles of Psychologists and Code of Conduct as set out by the British Association for Behavioural and Cognitive Psychotherapies and British Psychological Society. Ethical approval was received from the Surrey Research Ethics Committee and the NHS Health Research Authority (ref. 20/PR/0291). Local approvals to conduct the research in five NHS sites across England were granted.

References

Anastasi, A., & Urbina, S. (1997). *Psychological Testing*. Prentice Hall/Pearson Education.

Barrowcliff, A. L., Oathamshaw, S. C., & Evans, C. (2018). Psychometric properties of the Clinical Outcome Routine Evaluation-Learning Disabilities 30-Item (CORE-LD30). *Journal of Intellectual disability Research*, *62*, 962–973.

Beck, A. T. (2002). Cognitive models of depression. *Clinical Advances in Cognitive Psychotherapy: Theory and Application*, *14*, 29–61.

Breen, J. (2017). *Adapting the GAD-7 and PHQ-9 clinical measures for people with learning disabilities*. Unpublished doctoral dissertation, Royal Holloway, University of London.

British Psychological Society (2015). Guidance on the Assessment and Diagnosis of Intellectual Disabilities in Adulthood. https://www.bps.org.uk/sites/bps.org.uk/files/Member%20Networks/Faculties/Intellectual%20Disabilities/Guidance%20on%20the%20Assessment%20and%20Diagnosis%20of%20Intellectual%20Disabilities%20in%20Adulthood.pdf

Chinn, D., Abraham, E., Burke, C., & Davies, J. (2014). IAPT and Learning Disabilities. Foundation for People with Learning Disabilities. http://www.learningdisabilities.org.uk/content/assets/pdf/publications/iapt-and-learning-disabilities-report.PDF

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*, 2nd edn. Hillsdale, NJ: Erlbaum.

Cuthill, F. M., Espie, C. A., & Cooper, S.-A. (2003). Development and psychometric properties of the Glasgow Depression Scale for people with a learning disability: individual and carer supplement versions. *British Journal of Psychiatry*, *182*, 347–353. https://doi.org/DOI:10.1192/bjp.182.4.347

Dagnan, D., Burke, C. K., Davies, J., & Chinn, D. (2015). Learning Disabilities: Positive Practice Guide. Improving Access to Psychological Therapies (IAPT). Foundation for People with Learning Disabilities.

Dagnan, D., Rodhouse, C., Thwaites, R., & Hatton, C. (2022). Improving Access to Psychological Therapies (IAPT) services outcomes for people with learning disabilities: national data 2012–2013 to 2019–2020. *the Cognitive Behaviour Therapist*, *15*, e4.

Dean, E. E., Shogren, K. A., Hagiwara, M., & Wehmeyer, M. L. (2018). How does employment influence health outcomes? A systematic review of the intellectual disability literature. *Journal of Vocational Rehabilitation*, *49*, 1–13.

Department of Health (1999). *A National Service Framework for Mental Health*. London: HMSO.

Department of Health (2001). *Valuing People: A New Strategy for Learning Disability in the 21st Century*. London: HMSO.

Department of Health (2008). *Improving Access to Psychological Therapies. Implementation Plan: National Guidelines For Regional Delivery*.

Department of Health (2012). *IAPT three-year report: the first million patients*. Crown Copyright. http://www.iapt.nhs.uk/silo/files/iapt-3-year-report.pdf

Fitzpatrick, R., Davey, C., Buxton, M. J., & Jones, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment*, *2*(14), i–iv, 1–74.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (5th edn). Boston, Mass.: Houghton Mifflin Company.

HM Government (2010). *Equality Act 2010*. http://www.legislation.gov.uk/ukpga/2010/15/section/4

Hughes-McCormack, L. A., Rydzewska, E., Henderson, A., MacIntyre, C., Rintoul, J., & Cooper, S. A. (2017). Prevalence of mental health conditions and relationship with general health in a whole-country population of people with intellectual disabilities compared with the general population. *BJPsych Open*, *3*, 243–248.

Jahoda, A., & Markova, I. (2004). Coping with social stigma: people with intellectual disabilities moving from institutions and family home. *Journal of Intellectual disability Research*, *48*, 719–729. https://doi.org/10.1111/j.1365-2788.2003.00561.x

Kiani, R., & Miller, H. (2010). Sensory impairment and intellectual disability. *Advances in Psychiatric Treatment*, *16*, 228–235. doi: 10.1192/apt.bp.108.005736

Kline, P. (2000). *Handbook of Psychological Testing*, 2nd edn. London: Routledge.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kooijmans, R., Mercera, G., Langdon, P. E., & Moonen, X. (2022). The adaptation of self-report measures to the needs of people with intellectual disabilities: a systematic review. *Clinical Psychology: Science and Practice*, *29*, 250.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine*, *16*, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lippold, T., & Burns, J. (2009). Social support and intellectual disabilities: a comparison between social networks of adults with intellectual disability and those with physical disability. *Journal of Intellectual Disability Research*, *53*, 463–473.

McDonnell, C. G., Boan, A. D., Bradley, C. C., Seay, K. D., Charles, J. M., & Carpenter, L. A. (2019). Child maltreatment in autism spectrum disorder and intellectual disability: results from a population-based sample. *Journal of Child Psychology and Psychiatry*, *60*, 576–584.

McGurk, K., & Skelly, A. (2014). A quick guide for psychologists working in learning disability on the available clinical outcome measures currently in use. *The Bulletin of the Faculty for People with Intellectual Disabilities*, *2*, 28–32.

MENCAP (2018). Stigma and Discrimination. https://www.mencap.org.uk/learning-disability-explained/research-and-statistics/stigma-discrimination

MENCAP (n.d.). How common is learning disability? https://www.mencap.org.uk/learning-disability-explained/research-and-statistics/how-common-learning-disability

Merrells, J., Buchanan, A., & Waters, R. (2018). The experience of social inclusion for people with intellectual disability within community recreational programs: a systematic review. *Journal of Intellectual & Developmental Disability*, *43*, 381–391.

Mindham, J., & Espie, C. A. (2003). Glasgow Anxiety Scale for people with an intellectual disability (GAS-ID): development and psychometric properties of a new measure for use with people with mild intellectual disability. *Journal of Intellectual disability Research*, *47*, 22–30. https://doi.org/10.1046/j.1365-2788.2003.00457.x

Moss, S., Emerson, E., Kiernan, C., Turner, S., Hatton, C., & Alborz, A. (2000). Psychiatric symptoms in adults with learning disability and challenging behaviour. *British Journal of Psychiatry*, *177*, 452–456. https://doi.org/DOI:10.1192/bjp.177.5.452

NDTi (2016). Green Light Toolkit: a guide to auditing and improving your mental health service so that it is effective in supporting people with autism and people with learning disabilities (version 2). https://www.ndti.org.uk/uploads/files/Green_Light_Toolkit_22_Nov_2013_final.pdf

NICE (2016). Mental health problems in people with learning disabilities: prevention, assessment and management. The National Institute for Health and Care Excellence, NICE Guide (March), 1–34.

Patel, M., Lee, J. Y., & Scior, K. (2023). Psychometric properties of measures designed to assess common mental health problems and wellbeing in adults with intellectual disabilities: a systematic review. *Journal of Intellectual disability Research*, *67*, 397–414.

Paterson, L., Mckenzie, K., & Lindsay, B. (2012). Stigma, social comparison and self-esteem in adults with an intellectual disability. *Journal of Applied Research in Intellectual Disabilities*, *25*, 166–176. https://doi.org/10.1111/j.1468-3148.2011.00651.x

Raczka, R., Theodore, K., & Williams, J. (2018). An initial validation of a new quality of life measure for adults with intellectual disability: the Mini-MANS-LD. *Journal of Intellectual Disabilities*. https://doi.org/10.1177/1744629518787895

Sáez-Suanes, G. P., García-Villamisar, D., del Pozo Armentia, A., & Dattilo, J. (2020). Emotional dysregulation and uncertainty intolerance as transdiagnostic mediators of anxiety in adults with autism spectrum disorders and intellectual disability. *Research in Developmental Disabilities*, *106*, 103784.

Shah, N., Cader, M., Andrews, B. et al. (2021). Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): performance in a clinical sample in relation to PHQ-9 and GAD-7. *Health Qual Life Outcomes*, *19*, 260. https://doi.org/10.1186/s12955-021-01882

Simões, C., & Santos, S. (2016). Comparing the quality of life of adults with and without intellectual disability. *Journal of Intellectual Disability Research*, *60*, 378–388.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Vlissides, N., Beail, N., Jackson, T., Williams, K., & Golding, L. (2017). Development and psychometric properties of the Psychological Therapies Outcome Scale – Intellectual Disabilities (PTOS-ID). *Journal of Intellectual Disability Research*, *61*, 549–559. https://doi.org/10.1111/jir.12361

Vlissides, N., Golding, L., & Beail, N. (2016). A systematic review of the outcome measures used in psychological therapies with adults with ID. In *Psychological Therapies and People Who Have Intellectual Disabilities* (pp. 166–203). British Psychological Society.

Whittaker, M. (2023). *Do specific adaptations need to be made to IAPT interventions to make them more accessible to patients with a learning disability? Doctoral dissertation*, University of East Anglia.

Wigham, S., Taylor, J. L., & Hatton, C. (2014). A prospective study of the relationship between adverse life events and trauma in adults with mild to moderate intellectual disabilities. *Journal of Intellectual Disability Research*, *58*, 1131–1140.