Testing, testing, one, two, three: service user evaluation of three standard measures of mental health and wellbeing in an online counselling and support service for children and young people

Aaron Sefi* [1] , Ian Frampton PhD [1]

[1] University of Exeter

*Corresponding author: aaron@kooth.com Kooth, The Epworth, 25 City Road, London EC1Y 1AA

Keywords: Routine Outcome Measures; Online Counselling; Children and Young People

Abstract

Assessment measures have been widely adopted into mental health settings incorporating idiographic (personalised) and nomothetic (standardised) approaches. Online mental health support services have developed significantly, but with very little testing of these measures.

This study assesses the relative merits of acceptability, applicability and appropriateness of assessment measures online for children and young people. Acceptability was determined by whether users elected to complete the measure, applicability by whether they understood the questions, and appropriateness by their experience of completing the measure.

Three measures: the Short Warwickshire Emotional and Mental Wellbeing Scale (SWEMWBS); the Young Person's CORE; and the Strengths and Difficulties Questionnaire (SDQ) were randomly offered to 7235 new registrations to assess relative acceptability of the measures, and asked two follow up questions to assess applicability and appropriateness. Chi-square tests of independence were calculated to compare response rates for each measure, followed by post hoc pairwise comparisons and effect sizes calculations.

There was a high level of acceptability of the measures, with statistically significant differences between the less accepted SDQ and more accepted SWEMWBS. For applicability, the SWEMWBS was less understood and relatable than the other assessment measures. All the measures demonstrated a tolerable appropriateness. All differences were indicated at a small effect size.

These findings support the acceptability, applicability and appropriateness of the use of assessment measures in an online setting. The SWEMWBS demonstrated surprisingly low applicability against the more clinical measures, suggesting further investigation into

how young people ascribe meaning to the measures, and their motivations for completing them.

Implications for Practice

- Online mental health services need to consider the appropriate use of nomothetic and idiographic measures
- Children and young people will complete measures outside of direct engagement in an online setting
- Learning more about what people ascribe to measures completion in terms of meaning and motivation, will provide greater understanding of their use

Implication for Policy

- Nomothetic assessments have a place in online mental health services, but more work needs to be done to understand how users engage with nomothetic outcome measures in online settings.

1. **Introduction**

1.1 Mental health prevalence and access to services

The prevalence of formally diagnosable mental health conditions globally is estimated at 792 million people,or approximately one in ten of the world's population living with a disorder (Ritchie and Roser, 2018). However this estimation is limited to those with a formal diagnosis. Broader definitions of mental health suggest that one in five children experience problems globally (Kieling et al., 2011) echoed by a recent large survey of young people conducted in the UK indicating the scale of mental health problems in

England is much higher than previous estimates, with two in five young people aged 10 to 14 scoring above clinical thresholds for emotional problems, conduct problems or hyperactivity (Deighton et al., 2019). With the impact of Covid-19, this is expected to surge, with 20% of all adults and 15% of children in the UK expected to need *additional* support for depression, anxiety and post traumatic stress disorders and other mental health difficulties in the years to come (O'Shea, 2020).

Against this backdrop, and despite much 'hidden data', National Health Service (NHS) Child and Adolescent Mental Health services (CAMHs) in the UK are struggling to meet demand (Dubicka and Bullock, 2017) with over one quarter of referrals into CAMHs not accepted in 2016-17 with significant regional variability (Frith, 2017). Only one in four children and young people diagnosed with a mental health disorder are seen by a specialist mental health service (NHS Digital, 2017). The number of young people without a formal clinical diagnosis not being seen in services will be much higher.

1.2 Digital support

The need for innovation in delivering services and support to young people in England was outlined in the government-led *Future in Mind* (NHS England, 2015) report with specific reference made to the role of digital technology in providing part of the solution. This has been echoed in further reports including the *NHS Long Term Plan* (NHS Providers, 2019) outlining the further use of '*digitally-enabled mental health care'*, with a commitment that digitally enabled models of therapy will be widely rolled out across mental health pathways by 2021.

Aligned with this policy shift, there has been successful scaling up of online service provision, with a significant recent increase in Digital Health Technologies (NICE, 2019) incorporating practitioner/patient interaction in web-based and mobile phone applications (Hollis et al., 2015).

The service used for this study, Kooth (www.kooth.com), has developed a successful model of online support for young people, currently commissioned by over two thirds of Clinical Commissioning Groups (CCGs) in England, with over 100,000 users in 2018 (XenZone, 2019).

1.3 Routine outcome measures and assessment

With increased interest and scrutiny in innovative approaches to service delivery, so the need has grown to demonstrate evidence of effectiveness and achievement of outcomes. Routine outcome measures (ROMs) and screening tools have been widely adopted into young people's mental health settings over the last thirty years. The use of standardised outcome measures became more formalised with their adoption within the Children and Young People's Improving Access to Psychological Therapy (CYP-IAPT) programme. This has enabled some 'real-world' data to be produced into the impact of CAMHs interventions on anxiety and depression (Edbrooke-Childs, Edridge, et al., 2018) (Edbrooke-Childs, Wolpert, Zamperoni, Napoleone, & Bear, 2018).

More recently there has been a drive towards ROMs that favour the idiographic (personalised) over the nomothetic (standardised) approach (Jacob, Edbrooke-Childs, Law, & Wolpert, 2017). These include in particular, the goal-based outcome measure (GBO), which has demonstrated particular favourability on behalf of practitioners and young people alike (CORC, 2020). The use of GBOs have extended beyond CAMHS, and their uptake in online mental health settings has been shown to be beneficial (Hanley, Ersahin, Sefi, & Hebron, 2017).

1.4 Rationale

There has been very little testing of the nomothetic approach in an online remote environment. Some investigation has reported the complexity of their use in recording outcomes online (Sefi & Hanley, 2012), but little is known about how young people may accept, understand and respond to measurement in this context. It is tempting for services to adopt standard measures designed for face-to-face clinical settings, especially those recommended in CYP-IAPT, and employ them in online settings. However, this presumes a direct translation of these measures into a remote, online setting. In the case of Kooth, young people are able to register for the service pseudonymously, giving a self-constructed user name to represent them in the service. This is radically different from the referral-based, professionally-led process by which young people access CAMHs or school-based counselling. Therefore this study focuses on this key research question:

Are there discernable differences in the acceptability, applicability and appropriateness between three selected measures completed at registration for an online counselling service for young people?

2. Method

2.1 Selection of measures

This type of study benefits from a Participatory Action Research model (Baum, MacDougall, & Smith, 2006), where it is important to work iteratively with the participants, in this case stakeholders in the partner organisation to ensure the assessment measures would work appropriately in the novel context. As such, the measures were selected based on iterative discussion between the online counselling service and a range of researchers, policy makers and advisors, predominantly through its Research Advisory Board. This, coupled with direct experience of the implementation of a range of measures (Sefi & Hanley, 2012) helped form a shortlist of three with distinct attributes:

Strengths and Difficulties Questionnaire (SDQ)

The SDQ is an emotional and behavioural screening questionnaire (Goodman, 1997), and a commonly used screening tool internationally within both clinical and community populations, translated into over 80 languages (Stolk, Kaplan, & Szwarc, 2017). There has been some comparison of the use of SDQ in a computerised, remote setting, reporting a higher satisfaction around usage over face to face settings (Truman et al., 2003).

The SDQ was selected for the trial because of its international reputation, and its clinical focus in counterbalance to the other two measures, which are more commonly used in non-clinical populations.

Young Person's Clinical Outcomes in Routine Evaluation (YP-CORE)

YP-CORE was designed primarily for use in school counselling service settings (Twigg et al., 2009) and has been used in evaluating the introduction of statutory counselling in schools in Wales (Hill et al., 2011). It has also been used in Randomised Control Trials (RCTs) of school-based counselling (Stafford et al., 2018). Its subscales focus less on specific diagnostics and more on generalisable issues, such as wellbeing and functioning. It has been shown to be sensitive to change, but the accepted validity of the methodology for reporting this change has been challenged in an online setting (Sefi & Hanley, 2012).

It was selected for the trial due to its recognised thresholds, cut-offs and normative data and its design for settings beyond the clinic in wider populations, and its applicability to humanistic and integrative counselling models.

Short Warwick - Edinburgh Mental Wellbeing Scale (SWEMWBS)

The short version of the Warwick-Edinburgh mental well-being scale is used to monitor mental wellbeing in the general population, in particular by the Office of National Statistics (Peasgood, Brazier, & Mukuria, 2014)It has been demonstrated to have adequate internal consistency and reliability, which has been shown through its use internationally (Haver, Akerjordet, Caputi, Magee, & Furunes, 2015). It has become a

popular choice for both large-scale social surveys and intervention studies, due to its similar results to its longer counterpart, but with its brevity presumed to make it more acceptable (Ng Fat, Scholes, Boniface, Mindell, & Stewart-Brown, 2017).

It was selected for this study due to the perceived acceptability of the measure, and as a 'wellbeing' counterbalance to the more diagnostic SDQ and wider mental health counselling measure of the YP-CORE.

2.2 Implementation of measures and safeguarding risk

Once the measures were agreed, their implementation was explored. As the goal was to implement the measure as early in the use of service as possible, it was agreed to include them in the registration process of the online service. The style and format of the measure implementation was developed with user experience and user interface designers and software engineers employed by the service provider. Their considerations included ensuring the design would encourage completion and make users curious, whilst being clear that the measure could be skipped without repercussion; ensuring the content was clear and fit for purpose, whilst maintaining the integrity of the question format from the face-to-face environment; and testing some different options to find the most likely to be completed. The layout was tailored to the amount of answers shown, with a click through style flow agreed as offering the easiest to process experience.

The measures were allocated randomly to each new registration to the online service for ten weeks between September and November 2019. Over the six-week period, 7235 young people who signed up for the service, and consented to share their date for

research purposes, were randomly offered one of the three measures at completion of registration. As this was an innovative implementation of measures, it was important to ensure safeguarding protocols of the partner organisation were implemented, to ensure any risk factors were managed safely. These were agreed in collaboration with the clinical governance team.

2.3 Data collection

As this study was focused on acceptability, applicability and appropriateness of each measure, data was systematically collected for each of these areas:

Acceptability

This involved capturing how many young people at registration chose to complete the measure they were assigned. They were given up to five days after registration to complete the measure, and up to three refusals, with logic built in to ensure the measure did not appear after third refusal.

Applicability

In order to ascertain whether the measure was applicable to young people, a follow up question was agreed upon: “Did you understand and relate to those questions?” With them able to answer yes or no, or skip the question.

Appropriateness

To gauge the level of appropriateness of each measure, a further follow up question was established: “How are you feeling after answering those questions?” With them able to select from the following answers: Better, Same, Worse, Unsure.

2.4 Data analysis

To analyse the statistical independence of each of the three described areas of investigation for the measures, chi square tests of independence were fitted to explore whether there were overall significant differences between the dichotomous dependent variables for each area. Once statistically significant differences in proportions were identified, these were followed up by multiple post-hoc pairwise tests, one for each of the pairwise comparisons, to establish the source of significant difference. With a significance level selected as $p < 0.05$, for the pairwise tests, the Bonferroni corrected alpha level was applied to establish the significance level at $p < 0.016$.

As the significant test does not tell us the degree of effect, effect sizes were also calculated to show the magnitude of effect. Effect sizes are important when using chi square tests (Cohen, 1988), as the confidence in association requires corroboration of the likely size or magnitude of the difference between predicted (random) and observed results. There is further outline on the analysis in the results section.

2.5 Ethical Approval

The study was approved by the University of Exeter Psychology Research Ethics Committee. Close consideration was paid to gaining consent for sharing anonymised data, which was sought at registration on the service. The data reflected in the results is from the 71% of the young people asked at registration who gave consent for their data to be used. All users who completed the measures and indicated a level of need were followed up by the partner organisation in line with their clinical governance process.

3. Results

3.1 Primary Findings

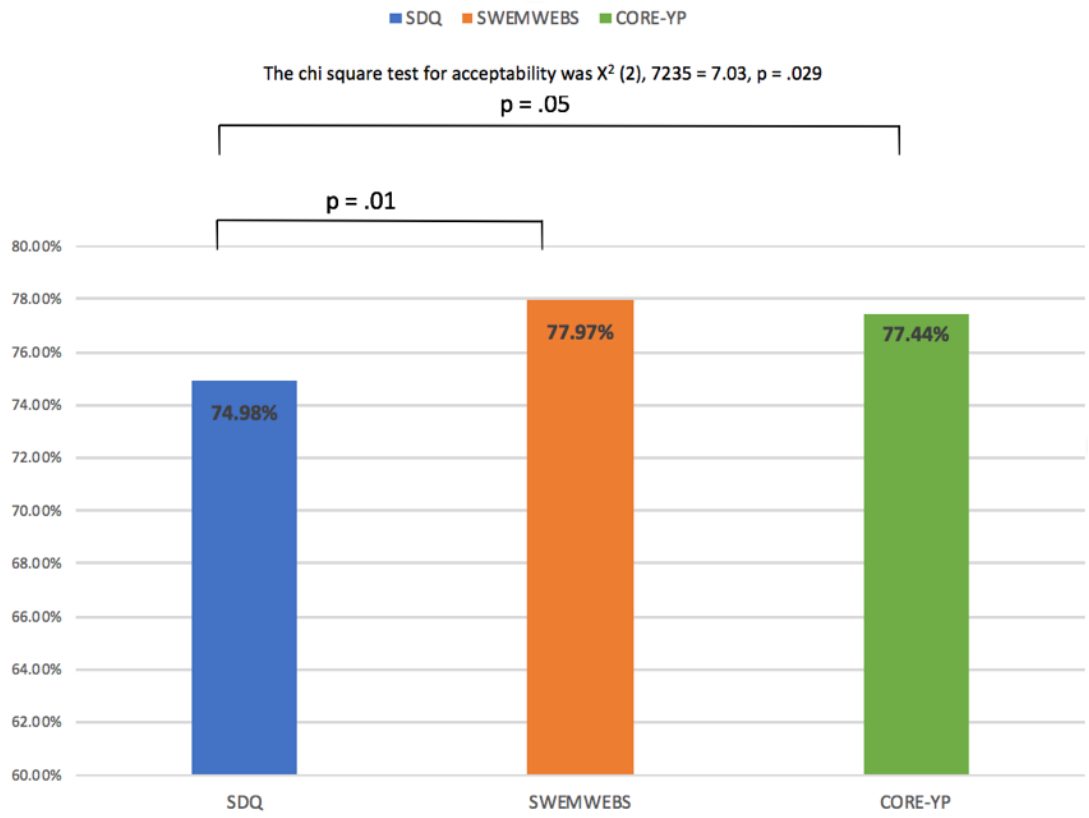

Figure 1: Acceptability - Completion Rates of Measures

For Acceptability, the relative completion rates were considered. The chi square test in overall difference in completion rates for the measures was X^2 (2, 7235) = 7.03, p = .029. The association was moderately strong. (Cohen, 1988). As indicated in Figure 1, in the post-hoc pairwise tests, differences were statistically significant between SWEMWBS and SDQ, with the significance level between CORE-YP and the SDQ not significant, when factoring in the Bonferroni correction, i.e. $p < 0.016$.

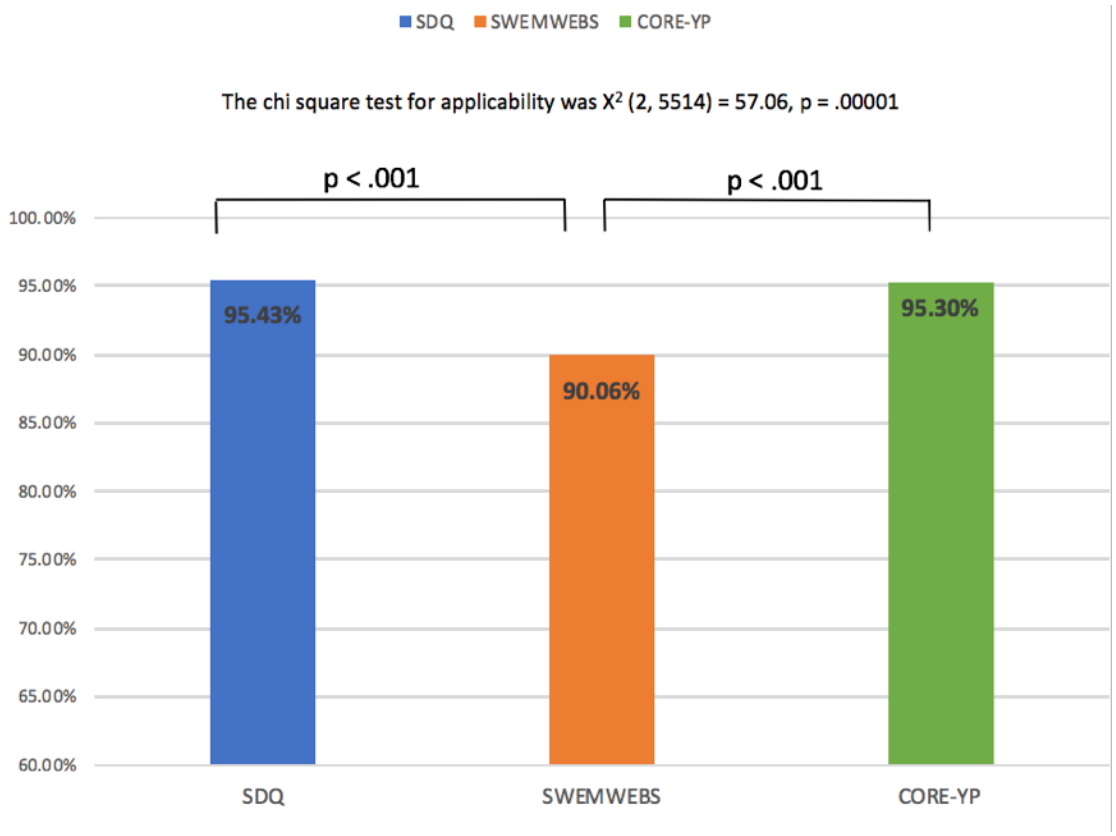

Figure 2: Applicability - "Did you understand and relate to the questions?"

For Applicability, the response to whether the user understood and related to the questions was considered. The chi square test for relative applicability of the measures was X^2 (2, 5514) = 57.06, p < .001, indicating a statistically significant association between measure and applicability. As indicated in Figure 2, in the post-hoc pairwise tests, differences were also statistically significant between SWEMWBS and SDQ, and SWEMWBS and CORE-YP respectively, following Bonferroni correction.

For Appropriateness, the analysis was broken down to compare the four different response options. For those who responded, “better” or “worse” there was no

statistically significant associations between response rate and measure. However, for those reporting “same” the chi square test for pairwise comparison between SDQ and SWEMWBS was X^2 (1, 3762) = 33.55, p = .00001. The chi square test for pairwise comparison between YP-CORE and SWEMWBS was X^2 (1, 3691) = 14.42, p = .0001. See Figure 3.

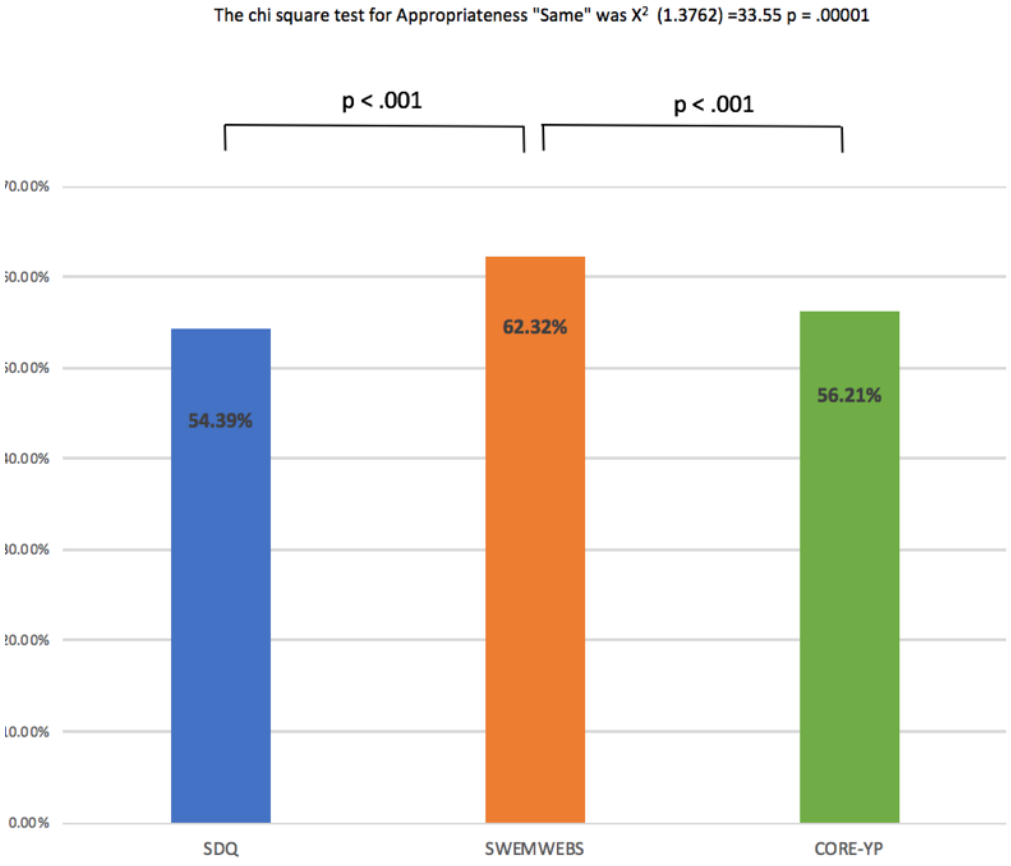

Figure 3: Appropriateness - How are you feeling? “Same”

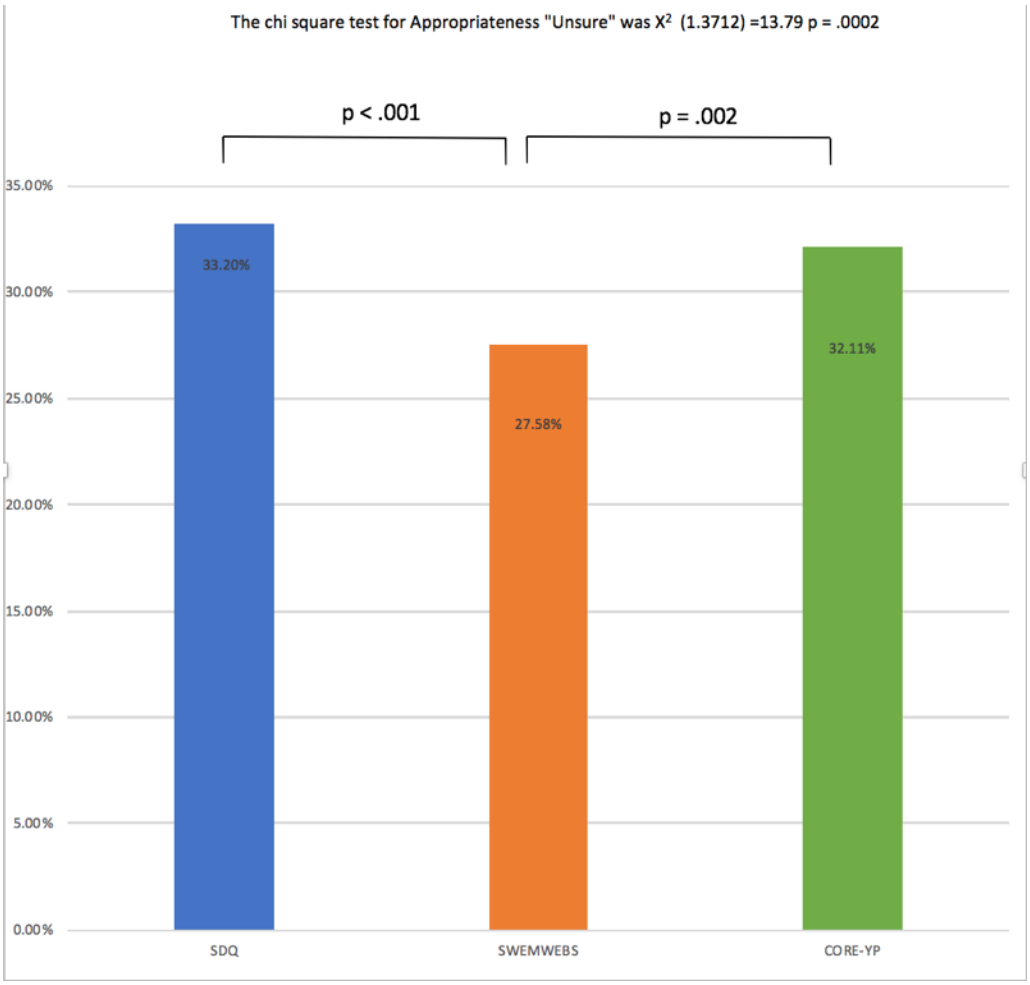

Figure 4: Appropriateness - How are you feeling? "Unsure"

For those reporting "unsure" the chi square test for pairwise comparison between SDQ and SWEMWBS was X^2 (1, 3712) = 13.79, p = .0002. The chi square test for pairwise comparison between YP-CORE and SWEMWBS was X^2 (1, 3691) = 9.13, p = .003. See Figure 4.

3.2 Effect Size

The significance test does not tell us the magnitude of the difference between the measures for acceptability, applicability and appropriateness. It is therefore helpful to show the magnitude of the effect size using Phi (φ) Square, applied to the pairwise comparisons in Table 1.

	Comparison		
	SDQ/SWEMWBS Phi	SDQ /CORE Phi	CORE/SWEMWBS Phi
Acceptability	0.04	0.03	0.01
Applicability	0.11	0.01	0.11
Appropriateness			
Better	*	*	*
Same	0.12	0.04	0.08
Worse	*	*	*
Unsure	0.11	0.02	0.09

Table 1: Effect size (Phi) for pairwise comparison between measures.

These findings all indicate a small or very small effect size of the magnitude of difference in the pairwise comparisons.

4. Discussion

4.1 Contextualising the findings

The measures used in this study were originally developed for specific contexts: identifying clinical 'caseness' (SDQ); evaluating large scale counselling interventions (YP-CORE); or wellbeing measurement (SWEMWBS). The need for self-report measures is apparent, especially considering the frequent divergence between clinician rating and self-report (Salmond, 2020). Yet to date there has been very little exploration

of the relative merits of their acceptability (do young people fill these out?), applicability (do they understand and relate to the questions?) and appropriateness (do they feel better, worse, the same, or even know how they feel as a result of completing the measures?).

4.2 Interpreting the findings

As descriptive statistics, the data indicate that overall, there is a high level of acceptability of all three measures in an online setting, with between 72-75% completion rate. Given the high emphasis placed on the choice to complete being made explicit at various points in the completion cycle, this represents a highly successful completion rate, and indicates that any of these measures would be acceptable in this setting. Perhaps this is not a surprisingly high rate given that 'digitally wise' natives (Prensky, 2009) are the participants of this study, but there is little or no existing evidence that standardised assessments would be accepted in this medium. In terms of applicability, again a high level of understanding and relatability of the measures was described, with between 90-95% reporting to the positive. Again, as a novel study, there are no benchmarks to compare this level of applicability, but it is clear that users will self-describe an understanding of these measures. Finally, appropriateness was indicated with the majority 54-62% saying they felt the same as a result of completing the measure, and 6-7% feeling worse, 4-5% feeling better and with many (27-33%) unsure. The level of 'unsure' suggested a lack of clarity around the question, with a clear sense that the majority did not report any significant change in mood after completing the measure.

The chi squares for goodness of fit (Cohen & Sackrowitz, 1975) each indicated there was some value in establishing post-hoc pairwise comparisons. In these comparisons, it was apparent that other than in the acceptability test, the SWEMWBS measure was consistently the outlier. The *acceptability* test showed the SDQ as the least completed measure, which was perhaps to be expected, given the relative length of this measure and its design as more clinical (with a common subsequent assumption as less acceptable) measure. However, the magnitude of difference here, indicated by the small effect size, was less than might have been expected. It is worth asserting the significance of this finding – that between the three measures, there is little difference in their generally high completion rates, and that the SDQ is, perhaps more surprisingly, acceptable to young people in this context. Whilst there has been very little research done on comparison between measures in this way, there has been some focus on 'idiographic' and user -friendly measures (Jacob et al., 2017) or what is most counted in mental health (Krause, Bear, Edbrooke-Childs, & Wolpert, 2019). These findings suggest that a more symptom-based measure is at least as acceptable as more wellbeing based measures, but this needs to be tested in more qualitative studies.

For the *applicability* test, we see slightly stronger association, with a marked difference between SWEMWBS and the other two measures. The variance here however goes against the presumed value of this measure for wide and general use, in that the SWEMWBS is noticeably *less* understandable and relatable than the other more mental health focussed measures, with the SDQ faring best of all in this regard, suggesting that

young people may find more mental health focussed questions more relatable and understandable than general wellbeing questions. Whilst there is a small effect size associated with this variance, it is nonetheless noteworthy and demanding of further investigation.

Finally, the *appropriateness* test has less discernible findings. Whilst there was some suggestion of variation between SWEMWBS and the other measures, with more recording feeling the "same" and less were "unsure", it is not clear what difference in meaning young people would have ascribed to these two responses. There is very little to compare against with these findings, with the bulk of research focussed on practitioner views and approaches to use of measures, with many reporting positive views (Hall et al., 2014), but also pointing to lack of training and understanding around the use of measures (Martin, Fishman, Baxter, & Ford, 2011).

4.3 Strengths and Limitations of the study and next steps

This naturalistic study was unique in its approach – examining the use of nomothetic assessments directly completed by users online, and capturing further valuable data on acceptability and applicability. It also benefitted from a very large data set, enabling some conclusions to be drawn on the relative acceptability of these measures. To this extent it helps to inform practice around the application of measures, particularly in the online mental health context. With some significant differences, but small effect sizes, it has modest but meaningful implications for service delivery. The study attempted to quantitatively explore these differences, using the chi square tests. In doing so, it raised

a lot of further questions about the process of completing measures in this setting. As the online setting predicates a remoteness, there is an opportunity to explore the motivations of young people in the completion of measures, and also what meaning they ascribe to their completion. A future qualitative study looking at these motivations and meaning making will give greater insight into the acceptability and applicability of the measures, providing a window into how young people view measures generally, and in particular in an online setting.

An important next step with the data from these measures is to analyse the assessment scores from the responses. These will give important insight into the relative assessed clinical needs and wellbeing levels of the population utilising online services. Looking at benchmarks from more formal clinical settings, and normative data from general populations, we will be able to distinguish the particular levels of need in online counselling services. Another important step will be to deploy one of the nomothetic measures as an outcome metric with administration at different time points in the user journey online, to examine standardised 'distance travelled' alongside the use of idiographic measures.

Other important considerations for measures deployment in an online setting include the timing of measure completion – i.e. given its opt-in nature, how long does it take users to elect to complete the measure, and crucially service usage patterns – i.e. exploring any relationship between completing a measure and the subsequent elected use of service.

All these developments will further enhance our understanding of what is happening when young people fill out measures online, and this in turn can impact on the integral validity of routine outcome monitoring, both online and in face-to-face settings.

References

Baum, F., MacDougall, C., & Smith, D. (2006). Participatory action research. *Journal of Epidemiology and Community Health*, *60*(10), 854–857. https://doi.org/10.1136/jech.2004.028662

Cohen, A., & Sackrowitz, H. B. (1975). Unbiasedness of the Chi-Square, Likelihood Ratio, and Other Goodness of Fit Tests for the Equal Cell Case. *The Annals of Statistics*, *3*(4), 959–964. https://doi.org/10.1214/aos/1176343197

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge. https://doi.org/https://doi.org/10.4324/9780203771587

Deighton, J., Lereya, S. T., Casey, P., Patalay, P., Humphrey, N., & Wolpert, M. (2019). Prevalence of mental health problems in schools: Poverty and other risk factors among 28 000 adolescents in England. *British Journal of Psychiatry*, *215*(3), 565–567. https://doi.org/10.1192/bjp.2019.19

Dubicka, B., & Bullock, T. (2017, September 29). Mental health services for children fail to meet soaring demand. *BMJ (Online)*. BMJ Publishing Group. https://doi.org/10.1136/bmj.j4254

Edbrooke-Childs, J., Edridge, C., Averill, P., Delane, L., Hollis, C., Carven, M., … Wolpert, M. (2018). A feasibility trial of Power Up: A smartphone app to support patient activation and shared decision making for mental health in young people

(Preprint). *JMIR MHealth and UHealth*, *7*(6), 1–15. https://doi.org/10.2196/11677

Edbrooke-Childs, J., Wolpert, M., Zamperoni, V., Napoleone, E., & Bear, H. (2018). Evaluation of reliable improvement rates in depression and anxiety at the end of treatment in adolescents. *BJPsych Open*, *4*(4), 250–255. https://doi.org/10.1192/bjo.2018.31

Frith, E. (2017). *Access and waiting times in children and young people's mental health*. Retrieved from https://epi.org.uk/wp-content/uploads/2018/01/EPI_Access-and-waiting-times_.pdf

Hall, C. L., Moldavsky, • M, Taylor, • J, Sayal, • K, Marriott, • M, Batty, • M J, … Batty, M. J. (2014). Implementation of routine outcome measurement in child and adolescent mental health services in the United Kingdom: a critical perspective. *Eur Child Adolesc Psychiatry*, *23*, 239–242. https://doi.org/10.1007/s00787-013-0454-2

Hanley, T., Ersahin, Z., Sefi, A., & Hebron, J. (2017). Comparing Online and Face-to-Face Student Counselling: What Therapeutic Goals Are Identified and What Are the Implications for Educational Providers? *Journal of Psychologists and Counsellors in Schools*, *27*(1), 37–54. https://doi.org/10.1017/jgc.2016.20

Haver, A., Akerjordet, K., Caputi, P., Magee, C., & Furunes, T. (2015). Measuring mental well-being: A validation of the Short Warwick–Edinburgh Mental Well-Being Scale in Norwegian and Swedish. *Scandinavian Journal of Public Health*, *43*(7), 721–727. https://doi.org/10.1177/1403494815588862

Hill, A., Cooper, M., Pybis Research Facilitator, J., Karen Cromarty, B., Pattison, S., Spong Claire Dowd, S., … Kate Smith, B. (2011). *2 Evaluation of the Welsh School-based Counselling Strategy: Final Report Evaluation Team*. Retrieved from

www.uk.sagepub.com

Hollis, C., Morriss, R., Martin, J., Amani, S., Cotton, R., Denis, M., & Lewis, S. (2015). Technological innovations in mental healthcare: harnessing the digital revolution. *The British Journal of Psychiatry : The Journal of Mental Science*, *206*(4), 263–265. https://doi.org/10.1192/bjp.bp.113.142612

Jacob, J., Edbrooke-Childs, J., Law, D., & Wolpert, M. (2017). Measuring what matters to patients: Using goal content to inform measure choice and development. *Clinical Child Psychology and Psychiatry*, *22*(2), 170–186. https://doi.org/10.1177/1359104515615642

Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigbodun, O., … Rahman, A. (2011, October 22). Child and adolescent mental health worldwide: Evidence for action. *The Lancet*. Elsevier. https://doi.org/10.1016/S0140-6736(11)60827-1

Krause, K. R., Bear, H. A., Edbrooke-Childs, J., & Wolpert, M. (2019). Review: What Outcomes Count? A Review of Outcomes Measured for Adolescent Depression Between 2007 and 2017. *Journal of the American Academy of Child and Adolescent Psychiatry*, *58*(1), 61–71. https://doi.org/10.1016/j.jaac.2018.07.893

Martin, A. M., Fishman, R., Baxter, L., & Ford, T. (2011). Practitioners' attitudes towards the use of standardized diagnostic assessment in routine practice: A qualitative study in two Child and Adolescent Mental Health Services. *Clinical Child Psychology and Psychiatry*, *16*(3), 407–420. https://doi.org/10.1177/1359104510366284

Ng Fat, L., Scholes, S., Boniface, S., Mindell, J., & Stewart-Brown, S. (2017). Evaluating

and establishing national norms for mental wellbeing using the short Warwick–Edinburgh Mental Well-being Scale (SWEMWBS): findings from the Health Survey for England. *Quality of Life Research*, *26*(5), 1129–1144. https://doi.org/10.1007/s11136-016-1454-8

NHS Digital. (2017). Mental Health of Children and Young People in England, 2017 [PAS] - NHS Digital. Retrieved February 19, 2020, from https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2017/2017

NHS England. (2015). *Future in mind - Promoting, protecting and improving our children and young people's mental health and wellbeing*. Retrieved from www.gov.uk/dh

NHS Providers. (2019). NHS Mental health implementation plan, (July), 7pp. Retrieved from https://nhsproviders.org/resource-library/the-mental-health-implementation-plan-201920-202324

NICE. (2019). Evidence standards framework for digital health technologies. *Nice*, (March), 1–35. https://doi.org/10.1007/978-1-349-95810-8_867

O'Shea, N. (2020). *Covid-19 and the nation's mental health*. Retrieved from https://www.centreformentalhealth.org.uk/sites/default/files/publication/download/CentreforMentalHealth_COVID_MH_Forecasting3_Oct20_0.pdf

Peasgood, T., Brazier, J., & Mukuria, C. (2014). Policy Research Unit in Economic Evaluation of Health and Social Care Interventions Research Report A conceptual comparison of well-being measures used in the UK Number September 2014, (September), 1–55.

Prensky, M., & Prensky, M. (2009). H. Sapiens Digital: From Digital Immigrants and

Digital Natives to Digital... *Innovate: Journal of Online Education*, *5*(3).

Ritchie, H and Roser, M. (2018). Mental Health - Our World in Data. Retrieved September 30, 2020, from https://ourworldindata.org/mental-health#citation

Salmond, C. H. (2020). Use of self-report outcome measures in an acute inpatient CAMHS setting. *Clinical Child Psychology and Psychiatry*, *25*(1), 174–188. https://doi.org/10.1177/1359104519869415

Sefi, A., & Hanley, T. (2012). Examining the complexities of measuring effectiveness of online counselling for young people using routine evaluation data. *Pastoral Care in Education*, *30*(1). https://doi.org/10.1080/02643944.2011.651224

Stafford, M. R., Cooper, M., Barkham, M., Beecham, J., Bower, P., Cromarty, K., … Street, C. (2018). Effectiveness and cost-effectiveness of humanistic counselling in schools for young people with emotional distress (ETHOS): study protocol for a randomised controlled trial. *Trials*, *19*(1), 175. https://doi.org/10.1186/s13063-018-2538-2

Stolk, Y., Kaplan, I., & Szwarc, J. (2017). Review of the strengths and difficulties questionnaire translated into languages spoken by children and adolescents of refugee background. *International Journal of Methods in Psychiatric Research*, *26*(4). https://doi.org/10.1002/mpr.1568

Truman, J., Robinson, K., Evans, A. L., Smith, D., Cunningham, L., Millward, R., & Minnis, H. (2003). The Strengths and Difficulties Questionnaire: A pilot study of a new computer version of the self-report scale. *European Child and Adolescent Psychiatry*, *12*(1), 9–14. https://doi.org/10.1007/s00787-003-0303-9

Twigg, E., Barkham, M., Bewick, B. M., Mulhern, B., Connell, J., & Cooper, M. (2009).

The young person's CORE: Development of a brief outcome measure for young people. *Counselling and Psychotherapy Research*, *9*(3), 160–168. https://doi.org/10.1080/14733140902979722

XenZone. (2019). *Kooth of Today: A Year in Review*. Retrieved from https://xenzone.com/wp-content/uploads/2019/04/Kooth-A-Year-In-Review.pdf

Aaron Sefi has evaluated school-based and online counselling services, and the use of appropriate outcome measures. He is currently overseeing the research strategy for XenZone, developing the evidence base for online therapeutic support and working with academic partners to advance the rigour around the evidence base for this medium. He is a frequent speaker on this, and the subject of defining appropriate outcome measures, and works closely with NHS England, Public Health England and other key bodies on policy developments in these areas. He provides guest lecturing on these subjects, as well as training, and has written numerous training sessions for the Counselling MindED e-learning programme. He lives by the sea in his native land Cornwall. ORCiD: https://orcid.org/0000-0002-0776-3858 @aaronsefi

Disclosure Statement: Aaron Sefi is an employee at XenZone, the platform utilised for the data collection for this research.