

Stigma and Health

Psychometric Validation of the Reported and Intended Behavior Scale (RIBS) with Adolescents --Manuscript Draft--

Manuscript Number:	SAH-2019-0038R1
Full Title:	Psychometric Validation of the Reported and Intended Behavior Scale (RIBS) with Adolescents
Abstract:	<p>Despite increasing interest in mental health education to reduce stigma, few studies assess changes in self-reported and intended discriminatory behavior. The current study evaluated the psychometric quality of the Reported and Intended Behavior Scale (RIBS) with adolescents. Participants were 11-15-year-olds from England (N = 1,032, 58% female). Confirmatory factor analysis (CFA) established a two-factor structure. The intended behavior scale showed high internal consistency ($\alpha = .94$, $\omega = .94$) and observed ceiling effects. A moderate correlation was found between intended behavior and stigma-related knowledge ($r = .39$). The average reading age was 14 years; however, the introductory text had a high reading age and might benefit from being simplified in future use. Females and early-adolescents (aged 11-13 years) reported more positive intended behaviors overall, with some group differences in item response. Multi-group CFA revealed partial scalar measurement invariance. Future research should assess self-reported and intended behavior and be cautious when investigating mean differences for gender and age.</p>
Article Type:	Article
Keywords:	stigma; discriminatory behavior; mental health; adolescence; validation
Corresponding Author:	Rosie Mansfield, MPhil The University of Manchester Manchester, UNITED KINGDOM
Corresponding Author E-Mail:	rosie.mansfield@postgrad.manchester.ac.uk
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	The University of Manchester
Other Authors:	Neil Humphrey, PhD Praveetha Patalay, PhD
Corresponding Author's Secondary Institution:	
First Author:	Rosie Mansfield, MPhil
Order of Authors Secondary Information:	
Manuscript Region of Origin:	UNITED KINGDOM
Suggested Reviewers:	<p>Luca Pingani, PhD Local Health Agency of Reggio Emilia luca.pingani@ausl.re.it Dr Luca Pingani published an Italian translation and validation of the Reported and Intended Behaviour Scale in 2016 with an adult sample.</p> <p>Sosei Yamaguchi, PhD National Center of Neurology and Psychiatry sosei.yama@ncnp.go.jp Dr Sosei Yamaguchi developed a Japanese version of the Reported and Intended Behaviour Scale in 2014, and assessed the reliability and validity of the scale.</p> <p>Vanessa Pinfold, PhD McPin Foundation contact@mcpin.org Dr Vanessa Pinfold has published research on stigma and discrimination, and has</p>

	<p>conducted evaluations of mental health education interventions in secondary schools.</p> <p>Melissa Pinto-Foltz, PhD Case Western Reserve University Melissa.Pinto-Foltz@case.edu Dr Melissa Pinto-Foltz has conducted evaluations of knowledge-contact interventions to reduce adolescent mental illness stigma. Melissa has also conducted psychometric validations of scales.</p>
Opposed Reviewers:	
Order of Authors:	<p>Rosie Mansfield, MPhil</p> <p>Neil Humphrey, PhD</p> <p>Praveetha Patalay, PhD</p>

Reviewer Comment	Author Response and Action
Reviewer 1	
<p>In abstract, please provide an explanation for what ages "early adolescents" reflects, or alternatively only report the relevant age-bracket.</p>	<p>We accept that this needed further clarification. The relevant age-bracket has now been added following the term "early adolescents" in the abstract. See page 1.</p>
<p>In the first paragraph of the introduction, might it be possible to source a more recent reference to strengthen the argument that youth populations are underrepresented in stigma research? I do not disagree with this point, but on its own the current 2004 reference is 15 years old, and given the proliferation of stigma research in the area it could be expected that more studies focusing on children and adolescents have since been produced.</p>	<p>A recent article by DeLuca (2019) has now been added to the first paragraph of the introduction. In this review of adolescent stigma research, the author also cites Link et al. (2004) when demonstrating the underrepresentation of children and young people in the field. Despite recognising that there is an increasing interest in understanding stigma in this population, they report that research that accounts for multiple-dimensions of stigma using reliable and valid measures is still limited. We feel that adding this recent review to the introduction provides a more up to date argument that adolescents are still unrepresented, and provides further rationale for the current study.</p>
<p>In the Design section, it would be good to also include the name of the GHSQ in relation to assessing construct validity to clarify how this assessment was done.</p>	<p>We agree that this needed clarifying. The following passage has now been added to the design section. See page 5.</p> <p>"Construct validity is assessed using a latent correlation with the General Help-Seeking Questionnaire (GHSQ)."</p>
<p>Can you please clarify whether the 35 students that opted out from the study via parental declined consent is an N or %? Is there information on how many, if any, pupils dropped out at the stage of providing assent? Is it possible to report on a "total number approached/considered for study" alongside the n=1132 sample that completed the survey?</p>	<p>The number of pupils opted out by parents has now been clarified on page 6 as N = 35. The following sentence has also been added on page 6 to provide information of pupils dropping out at the stage of assent:</p> <p>"Seven pupils were removed that had not provided assent leaving a sample of N = 1,132 pupils that assented to completing the survey."</p> <p>The total number of participants recruited has been added (N = 1,823). The number of schools lost from recruitment to follow up (N = 3) has also been noted on page 6 and the final sample for analysis (N = 1,032) calculated as a % of the initial number of pupils considered.</p>
<p>Could it please be specified exactly what demographic information were collected regarding participants (e.g. at the beginning of</p>	<p>The following sentence has been added to the measures section to clarify what demographic information was collected from participants:</p>

<p>the methods section; currently this is more implied in reporting data later)?</p>	<p>“Gender and year group was the only demographic information collected from pupils in the current study.”</p>
<p>If n=100 students' data was removed due to missing all RIBS items, what proportion was missing data for some RIBS items? What procedure was followed for these persons?</p>	<p>This information is already provided on page 6 when reporting the MCAR analysis. There was a range of missingness at an item level (.6-2.1%), however, given that missingness was <.5% no imputation methods were used in the current study. The following sentence has been added to help clarify this.</p> <p>“Due to low levels of item level missing data, no imputation methods were utilized in the current study; the WLSMV estimator uses all available data using pairwise present.”</p>
<p>It is also great to see that the level of missingness was examined, e.g. using MCAR. But did these examinations of missing data lead to e.g. using multiple imputation or other methods to account for the missing data?</p>	<p>We feel that the response to the previous comment also answers this query.</p>
<p>The abbreviation for SEM is explained twice in the manuscript. The abbreviation WLSMV is currently used once before the explanation for it is later provided on p.8.</p>	<p>The abbreviation for structural equation modelling (SEM) is now defined on page 6 and then used on page 9. Similarly, WLSMV is now defined on page 6 and the abbreviation used from then on.</p>
<p>(p.6, measures) I would recall that the explanation for the RIBS explains that people with mental health problems could be "for example/e.g." people seen by healthcare staff, not that this is the only kind of population that is of interest. So this would seem more like "exemplifying" than "defining" the sample in these terms.</p>	<p>We agree that “exemplifying” is a more appropriate choice of wording here and have adapted the measures section accordingly.</p>
<p>Across the result and discussion, when discussing the RIBS scores, the authors speak of higher scores as indicating "more positive attitudes". Given that the RIBS is a measure for discrimination/behaviour specifically, in order for a reader to not assume this is another measure assessing attitudes, I would recommend that the authors rephrase this as e.g. "more intended behaviours/social contact", or "more intended contact; i.e. less intended discriminatory behaviours". Or at least when referring to the score speak of this in terms of "intended behaviours" rather than "attitudes", given the nature of the scale.</p>	<p>We agree that “more positive attitudes” could cause confusion so have changed the wording to “positive intended behaviours” throughout the manuscript.</p>

<p>When discussing the MAKS (p.7), can the authors please specify whether a total score was used?</p>	<p>This is clarified on pages 7-8 with the following sentence:</p> <p>“In the current study a total score was calculated (6-30), with a higher score indicating better stigma-related knowledge.”</p>
<p>Can psychometric properties of MAKS be reported (p.7), as this is done for the RIBS and the GHSQ.</p>	<p>The psychometric properties of MAKS have now been added for the original adult sample and an adolescent sample. See the following sentence on page 8:</p> <p>“The scale (items 1-6) was found to have moderate internal consistency (Cronbach’s alpha = .65) with an adult sample (Evans-Lacko et al., 2010) and low internal consistency (Cronbach’s alpha = .24) in an adolescent sample (Chisholm et al., 2016).”</p> <p>Although the MAKS shows low internal reliability in this previous study, we report moderate levels of internal consistency in the current study (see table 2). We also acknowledge the limitations of using MAKS with this population in the discussion.</p>
<p>Regarding GHSQ, in the analyses was the averaged mean likelihood score used (p.8), or were the scores for each item considered separately (as implied in the top of p.9)?</p>	<p>We agree that the wording on page 9 might be misleading. The wording on page 9 has now been updated to match page 8 i.e. “mean likelihood of help-seeking calculated from the GHSQ “</p>
<p>Can it be spelled out what MIMIC (p.9) stands for?</p>	<p>The following has been added to page 9 to spell out what MIMIC stands for:</p> <p>“Specifically, differential item functioning (DIF) comparing gender and age group was examined in a multiple indicator multiple cause (MIMIC) model.”</p>
<p>Should the "Readability and Validity" subheading be rephrased as "Validity" (given the next subheading is "readability")?</p>	<p>We checked the sub-heading titles and they read “Reliability and Validity” and “Readability” and have therefore not been changed.</p>
<p>On p.11; the association of 0.30 would seem "small to moderate"; I believe that by usual convention 0.5 would be considered moderate?</p>	<p>When presenting r as opposed to d, >.30 can be seen as moderate. Furthermore, the exact r values for the bivariate and latent correlations are presented in table 2 and are closer to .40. We therefore believe our interpretation of the effect size is in line with recommendations in the literature (now on page 12).</p>

<p>A minor typo on p.14; reference in first paragraph (For example, schizophrenia is more likely...." should read Evans-Lacko; currently Evans-lacko.</p>	<p>We have gone through the whole manuscript to ensure that any instances of Evans-lacko are changed to Evans-Lacko as suggested.</p>
<p>On p.17 an attempt to "qualitatively assess the interpretability of RIBS" is mentioned - I'm sorry if I missed this; which aspect of the manuscript is referred to here?</p>	<p>We agree that the choice of the word 'qualitative' here was unintentionally misleading. The phrase "qualitatively assess the interpretability of RIBS" was referring to an attempt to qualitatively discuss the readability assessments in the discussion, for example, difficult words identified. We have now removed this phrase and instead added the following sentence to suggest future qualitative work (page 18):</p> <p>"Future research should use qualitative approaches such as focus groups and cognitive interviewing to explore adolescents' experiences and interpretation of RIBS, to reveal any issues with readability and inform future revisions to the scale. For example, accounting for developmental stage by changing "work" to "school" in items 2 and 6."</p>
<p>Relatedly, I would have recommended that the authors include reflections on how qualitative approaches could be used in further research to examine the RIBS amongst adolescents, e.g. focus groups with young people to consider issues around readability and comprehension, and suggest appropriate revisions to the phrasing.</p>	<p>We feel that the addition of the sentence above on page 18 also addresses this suggested change.</p>
<p>It is great to see on p.17 that it is reflected on that whether the MAKS is suitable for adolescents might have impacted on the results. Can a similar consideration be extended for the GHSQ; has this been assessed amongst adolescent samples?</p>	<p>Thank you for this suggestion. Given that there has been some previous psychometric work conducted with adolescents (Wilson et al. 2005), and the high level of internal consistency found in the current study (>.85), we do not feel that the same considerations need to be made in the limitation section (now on page 18).</p>
<p>Reviewer 2</p>	
<p>Given the aim is "to evaluate the quality of the RIBS for measuring reported and future discriminatory behaviors towards individuals experiencing mental health difficulties in an adolescent sample" (p.4), a primary question emerges as to the appropriateness of the content of the RIBS for an older child/adolescent audience given the participants' ages (11-15 years), in addition to its psychometric properties. The authors</p>	<p>We accept that the wording in the discussion is unintentionally misleading and indicates that qualitative approaches were used to explore adolescents' experiences and interpretation of RIBS. We can confirm that in the current study, qualitative interviewing was not used to determine if adolescents' interpretations of the scale content were valid for its intended purpose. We have included suggestions for</p>

acknowledge the importance of developmental considerations in scale and item construction: "As a measure intended for adult readers, there is no evidence that adolescents are able to access the content of the scale" (p.4). The authors address this point primarily through a metric of readability (see p.4 "An assessment of the readability and measurement invariance for different age groups is therefore necessary"). Readability is certainly an important metric of appropriateness. However, appropriateness of the content in terms of developmental stage is also crucial - for example, changing "work" to "school"; considering if adolescents know what is meant by defining mental health problems as a "compromise between mental illness/mental distress" (p.5). Were interviews conducted to determine if adolescents' interpretations of the scale content were valid for its intended purpose? Why did the authors not consider providing different, developmentally-appropriate wording? Taken together with information on the limited readability of the RIBS for this age range, it questions the appropriateness of administering this scale.

future qualitative work with the addition of the following sentence:

"Future research should use qualitative approaches such as focus groups and cognitive interviewing to explore adolescents' experiences and interpretation of RIBS, to reveal any issues with readability and inform future revisions to the scale. For example, accounting for developmental stage by changing "work" to "school" in items 2 and 6."

In response to your comment relating to considering if adolescents know what is meant by defining mental health problems as a "compromise between mental illness/mental distress". This sentence is not actually presented to participants in the instructions, it is included by the authors of the scale as justification for using the term 'mental health problem' throughout the scale. I do agree that it would be interesting for future research to explore adolescents' interpretations of different mental health terminology e.g. mental illness vs. mental health problem vs. mental distress. In the discussion, research conducted by Leighton (2009) is noted relating to inconsistencies in young people's understanding of mental health related terms.

Given that we did not conduct qualitative interviews we did not provide different, developmentally appropriate wording. For the purposes of exploring the psychometric quality of RIBS with an adolescent population, the original wording was maintained given that this version is already being used in some adolescent research (e.g. Chisholm et al. 2016). We felt it was important to first explore the psychometric quality and readability of the original version of RIBS, and make suggestions for future work to develop a scale that is more appropriate. That said, despite high readability scores for the instructions and some items, from a psychometric perspective the scale functions well with high factor loadings and internal consistency. As mentioned in the discussion, readability assessments taken alone can lack reliability, particularly when applied to short extracts of text like items.

<p>It is unclear what is intended by this text: "The average reading age was 14 years; however, the introductory text was not appropriate for this age group". I presume that the intro text in question isn't appropriate in terms of being too high a reading age, as opposed to some other factor that would make it inappropriate?</p>	<p>The reading age for the instructions has now been added to the abstract to add clarity:</p> <p>"The average reading age was 14 years; however, the introductory text had a high reading age and might benefit from being simplified in future use."</p>
<p>The first few paragraphs of the introduction seemed to jump around in topics. In particular, the opening paragraph was not clearly related to the topic and did not include a justification for the current study. I would suggest that the information on top paragraph of p.3 better introduces and justifies the topic.</p> <p>There is clarity needed around the use of the term 'actual discrimination' or 'actual discriminatory behaviour.' The RIBS is a self-report measure, so it is not a measure of 'actual discrimination', but rather a measure of self-reported discrimination. I suggest reviewing the introduction and all other sections to achieve this clarity.</p> <p>Why was the 11-15 year old age group chosen? The authors state that "adolescence" is their target developmental period; this extends beyond 15 years and warrants discussion.</p>	<p>The first paragraph aims to introduce the rationale for the study in terms of the underrepresented population, the importance of better understanding adolescent stigma and the need to consider the multiple dimensions of stigma when conducting research. We have now also added a recent review by DeLuca (2019) which further highlights the need for assessments of the reliability and validity of measures for this population in order for researchers to cover all dimensions of stigma in their evaluations.</p> <p>Given that the multi-faceted nature of the stigma construct is introduced in the first paragraph, the second paragraph aims to outline existing conceptualisations of stigma. The third paragraph then focuses on the importance of researching discriminatory behaviour as a dimension of stigma and then the fourth provides evidence that this is an under researched dimension.</p> <p>For the reasons above we would like to keep the existing structure in the introduction.</p> <p>We did however agree that it should be made clearer that RIBS is measuring self-reported discriminatory behaviour and so we have applied this change throughout the manuscript.</p> <p>The age group is based on the secondary school age in England and corresponds to the fact that the data are drawn from two school-based trials in this age-group. We feel that we have made it clear throughout the manuscript that the study is focusing on 'early-adolescents' and mid-adolescents' however, we agree that the word 'convenience' should be included when describing the sample and this has been added to page 6.</p>

	<p>We have also added a sentence to the discussion that recognises this as a limitation of the current sample age range and suggests future research should expand the sample:</p> <p>“Furthermore, the convenience sample was limited to early and mid-adolescents, therefore future research should consider expanding the sample to include adolescents aged 16 years and older and compare item functioning.”</p>
<p>Why were floor/ceiling effects considered at a cut off of 15% of responses? Is there a reference or standard that was used to guide this decision?</p> <p>With respect to readability, did the authors consider re-running the analyses with the items with higher reading ages (e.g., > 12 years) excluded, to be sure that the results still held in terms of factor loadings, developmental trends, and differences between reported & intended discrimination? If so, what happened? If not, then please justify your decision.</p>	<p>We can see why this cut off may have appeared arbitrary. The 15% cut off is a quality criteria set by Terwee et al. 2007. The reference has now been added so that the sentence reads:</p> <p>“Floor and ceiling effects are reported if 15% or more of the sample receive the lowest or highest possible score (Terwee et al., 2007).”</p> <p>Given that the scale shows the original two-factor structure, and that the intended behaviour scale shows high factor loadings and internal reliability, it suggests that despite high readability scores for the instructions and some items, the scale appears to be functioning as it should be in this current sample. For the intended behaviour scale, Cronbach’s alpha was calculated with each item removed in turn and this did not show improvements in internal consistency. Due to the already very high level of internal consistency, these results were not included in the final manuscript. The fact that the scale appears to psychometrically be functioning quite well, perhaps indicates an issue with judging the age appropriateness of the scale using only readability assessments. We have therefore decided to unpick this a little more in the discussion when commenting on possible limitations of the readability assessments.</p>
<p>The serious limitations of using this scale due to its poor readability for this age range need to be thoroughly considered. As it stands, the discussion expresses some hesitation, but could go further. The scores in Table 3 generally do not support the use of the RIBS in younger adolescents. All readability scores were above the youngest participant (of 11.00 years), with only 2 items in the 11-year old range (11.12 and</p>	<p>Similar to the point made above, given the results from the psychometric assessment of the scale, readability results are taken with some caution. We have adapted the discussion to more thoroughly consider the limitations of using this scale given the readability assessments while also recognising that given the psychometric qualities of the scale, the readability assessments should not be taken</p>

11.21) and 2 in the 12-year old range (both 12.72). Other items range from 13.01-16.91 years. In particular, given that the instructions were at a reading age in excess of the oldest participant (at 19.55), there is limited evidence that the instructions to complete the scale were truly understood by participants. This then questions the validity of the ensuing ratings and scores.

alone when determining the reliability and validity of the scale.

Dear Editor,

We would like to submit our paper “Psychometric Validation of the Reported and Intended Behavior Scale (RIBS) with Adolescents” to be considered for publication in *Stigma and Health*.

Few studies assess actual as well as hypothetical discriminatory behaviour towards individuals experiencing mental health difficulties. The RIBS was developed for use with adults and has since been translated into Japanese and Italian and validated with adult samples. Despite the use of RIBS in school-based, stigma research, there is currently no study assessing its psychometric quality for use with adolescent samples. The current study aimed to assess the internal consistency, floor and ceiling effects, construct, convergent and content validity, and interpretability of RIBS to inform its future use in adolescent stigma research.

Individuals experiencing mental health difficulties consistently report experiences of discrimination which can lead to reduced social networks and structural inequalities. Evidence suggests that supportive social relationships prevent adolescent mental health difficulties and improve symptoms. Given that adolescence is increasingly prioritised as an important developmental phase for promoting positive mental health, we believe that valid measures of discriminatory behaviour are needed in order to understand ways to reduce it. Furthermore, the assessment of measurement invariance in this study adds to the literature on gender and age differences in stigmatising attitudes and behaviours. It also recognises the importance of exploring the psychometric properties of stigma constructs across groups before testing for mean differences.

We believe that this article falls within the scope of *Stigma and Health* given that it has implications for the future measurement of discriminatory behaviours in adolescent, mental health stigma research. The study utilises a large adolescent sample and adopts sophisticated methods to assess the psychometric properties of RIBS. With an increasing overlap between education and mental health policy, this article will be of interest to researchers across disciplines with the aim of reducing stigma and promoting child and adolescent mental health.

We can confirm that this article has not been submitted or published anywhere else, and hope you will consider this article for publication in *Stigma and Health*. We look forward to receiving your feedback.

Yours faithfully,

Rosie Mansfield, Neil Humphrey and Praveetha Patalay

Correspondence to:

Rosie Mansfield, Phone: +161 275 8290, Email: rosie.mansfield@postgrad.manchester.ac.uk

B4.8 Ellen Wilkinson Building,
Institute of Education, University of Manchester,
Oxford Road, Manchester, UK,
M13 9PL

Running Title: Validation of RIBS with adolescents

**Psychometric Validation of the Reported and Intended Behavior Scale (RIBS) with
Adolescents**

Rosie Mansfield^{a} MPhil, Neil Humphrey^a PhD and Praveetha Patalay^b PhD*

^a Institute of Education, University of Manchester

*^b Institute of Education and Faculty of Population Health Sciences, University College
London*

*Corresponding author: Rosie Mansfield, +44161 275 8290,

rosie.mansfield@postgrad.manchester.ac.uk, B4.8 Ellen Wilkinson Building, University of
Manchester, M13 9PL

Neil Humphrey, +44161 275 3404, neil.humphrey@manchester.ac.uk, A1.19 Ellen
Wilkinson Building, University of Manchester, M13 9PL

Praveetha Patalay, +44207 612 6051, p.patalay@ucl.ac.uk, Room 301 55-59 Gordon Square,
London, WC1H 0NU

Rosie Mansfield Bio

Rosie Mansfield is a PhD student in the Institute of Education, University of Manchester. She is also a research assistant on the Education for Wellbeing Programme, funded by the Department for Education. Her research interests include theory, methods and practice relating to school-based, mental health education.

Neil Humphrey Bio

Neil Humphrey is Professor of Psychology of Education and Head of the Manchester Institute of Education, University of Manchester. His research interests include mental health, social and emotional learning, and special educational needs (particularly autism spectrum conditions). Neil's research has been funded by the National Institute for Health Research,

Running Title: Validation of RIBS with adolescents

Education Endowment Foundation, and Department for Education, among others. He is the author of *Social and emotional learning: a critical appraisal*, published by Sage.

Praveetha Patalay Bio

Praveetha Patalay is an Associate Professor at UCL, based across the Centre for Longitudinal Studies and the MRC Unit for Lifelong Health and Ageing. Her research focuses on the development, risk/protective factors affecting mental health, the consequences of mental ill-health and the ways in which we can reduce the stigma surrounding mental illness and promote wellbeing. She uses large national studies, including the British Birth Cohorts, to investigate mental health through the life course, and works on evaluations of school-based intervention programmes to support and promote young people's mental health.

Acknowledgements: The data used in this study was collected as part of the Education for Wellbeing Programme funded by the Department for Education (grant number: EOR/SBU/2017/015). The programme also funds Rosie Mansfield's PhD. The views expressed in this article are those of the author(s) and not necessarily those of the Department for Education or its arm's length bodies, or other Government Departments.

The authors declare no conflict of interest and take sole responsibility for the content of this article. The authors would like to thank the other members of the Education for Wellbeing research team: Jessica Deighton PhD, Daniel Hayes PhD, Emily Stapley PhD and Anna Moore MSc from the Anna Freud National Centre for Children and Families Evidence Based Practice Unit, Emma Ashworth PhD and Joao Santos PhD from the Institute of Education, University of Manchester, Jan Boehnke from Dundee Centre for Health and Related Research and Eva-Maria Bonin from the London School of Economics and Political Science.

17 pages including introduction, method, results and discussion

Running Title: Validation of RIBS with adolescents

48 references

1 figure

4 tables

Running Title: Validation of RIBS with adolescents

Abstract

Despite increasing interest in mental health education to reduce stigma, few studies assess changes in self-reported and intended discriminatory behavior. The current study evaluated the psychometric quality of the Reported and Intended Behavior Scale (RIBS) with adolescents. Participants were 11-15-year-olds from England (N = 1,032, 58% female). Confirmatory factor analysis (CFA) established a two-factor structure. The intended behavior scale showed high internal consistency ($\alpha = .94$, $\omega = .94$) and observed ceiling effects. A moderate correlation was found between intended behavior and stigma-related knowledge ($r = .39$). The average reading age was 14 years; however, the introductory text had a high reading age and might benefit from being simplified in future use. Females and early-adolescents (aged 11-13 years) reported more positive intended behaviors overall, with some group differences in item response. Multi-group CFA revealed partial scalar measurement invariance. Future research should assess self-reported and intended behavior and be cautious when investigating mean differences for gender and age.

Key words: stigma, discriminatory behavior, mental health, adolescence, validation

Public Significance Statement: The Reported and Intended Behavior Scale (RIBS) is a valid measure of reported and intended discriminatory behavior for an adolescent sample.

However, caution should be taken when using the scale with adolescents under the age of 14 with regards to its readability. Studies assessing the effect of mental health education on stigma reduction should assess self-reported and intended discriminatory behavior, and be cautious when exploring mean differences for gender and age group.

**Psychometric Validation of the Reported and Intended Behavior Scale (RIBS) with
Adolescents – re-submitted [28/08/2019]**

Adolescent populations have been majorly underrepresented in the mental health stigma literature over the past few decades (Link, Yang, Phelan, & Collins, 2004). Understanding the extent of adolescent stigma is important, not least because stigmatizing attitudes are repeatedly found to predict the help-seeking intentions of young people (Clement et al., 2015; Gulliver, Griffiths, & Christensen, 2010). Despite a growing interest in adolescent stigma, and an increased number of evidence-based stigma reduction interventions, there remains a lack of reliable and valid measures for this population. Furthermore, few researchers are considering the multi-faceted nature of stigma when selecting the most appropriate measures for a given research question (DeLuca, 2019).

Existing conceptualizations of stigma suggest it is a process involving labelling, stereotyping, cognitive separation, emotional reactions, status loss, and discrimination (Corrigan, 2000; Link et al., 2004). Thornicroft, Rose and Kassam (2007) present it as a problem of “*knowledge (ignorance), attitudes (prejudice), and behavior (discrimination)*” (p.192). This conceptualization maps onto the stigma process such that ignorance increases stereotype beliefs which, in turn, cause negative attitudes, emotional reactions and possible discriminatory behavior. Stigma also exists at both a public and personal level, where perceived public and personal stigma are theorized as distinct constructs (Griffiths, Christensen, & Jorm, 2008). An individual can hold stigmatizing beliefs, perceive public stigma, and internalize experiences of stigma and discrimination, reducing the likelihood of help-seeking behaviors and disclosures (Thornicroft, 2008).

Perceptions of public stigma and experiences of discrimination are consistently reported by individuals experiencing mental health difficulties, leading to reduced social connections, as

well as structural inequalities in income, housing, and employment (Stuart, Arboleda-Florez, & Sartorius, 2012). Furthermore, in a study of adolescent stigma experiences, young people who had previously received a psychiatric diagnosis reported a high level of social stigma, including loss of friendships, family stigma, and discriminatory behavior from school staff (Moses, 2010). Supportive social networks and positive peer relations have been found to reduce the risk of adolescents experiencing mental distress and improve symptoms (Ciarrochi, Morin, Sahdra, Litalien, & Parker, 2017; Crush et al., 2018; Van Droogenbroeck, Spruyt, & Keppens, 2018). These findings highlight the importance of developing better methods of measuring discriminatory behaviors in adolescent populations, in addition to knowledge and attitudes, in order to understand the processes by which they can be reduced. Given the evidence that intended or hypothetical behavior does not always translate into actual behavior change (Eisenberg, Speer, & Hunt, 2012; Thornicroft et al., 2007), measuring self-reported as well as intended behavior towards individuals experiencing mental health difficulties is recommended when evaluating anti-stigma interventions.

Reviews of the stigma literature have revealed that the measurement of self-reported discriminatory behavior is limited and that hypothetical or intended behaviors are measured less frequently than attitudes (Corrigan, Morris, Michaels, Rafacz, & Rusch, 2012; Fox, Earnshaw, Taverna, & Vogt, 2017). One hundred and one scales were identified in a recent review of stigma measurement tools. However, only eight were validated for child and adolescent samples, and none measured discriminatory behavior (Wei, McGrath, Hayden, & Kutcher, 2017). In a review of school-based interventions with the aim of improving mental health knowledge, stigma and help-seeking behaviors, attitudes towards mental illness were found to be the most commonly measured outcome. Of the 21 studies that measured stigma, only six used validated scales; none captured self-reported discriminatory behaviors (Wei, Hayden, Kutcher, Zygmunt, & McGrath, 2013).

The lack of validated adolescent measures has led to the use of adult scales without consideration of their appropriateness or validity in this younger age-group. For example, in a recent evaluation of a school-based knowledge-contact intervention to reduce adolescent stigma, validated adult measures of mental health stigma and knowledge were utilized (Chisholm et al., 2016). One of these, the Reported and Intended Behavior Scale (RIBS) (Evans-Lacko et al., 2011) was developed and validated with a UK adult sample with the aim of capturing self-reported as well as intentional behaviors towards individuals experiencing mental health difficulties. The RIBS has since been translated into Japanese and Italian and validated with university students and adults (Pingani et al., 2016; Yamaguchi, Koike, Watanabe, & Ando, 2014). However, to date, the scale has not been validated with an adolescent sample.

As a measure intended for adult readers, there is no evidence that adolescents are able to access the content of the scale. There are also complex findings relating to age as a predictor of stigmatizing attitudes. Previous research found that although social distance and the belief that mental health difficulties were a sign of weakness *decreased* with age, perceived stigma, reluctance to disclose, and the belief that mental health difficulties made a person dangerous and unpredictable *increased* (Jorm & Wright, 2008). An assessment of the readability and measurement invariance for different age groups is therefore necessary. Furthermore, given that studies commonly find that adolescent males report more stigmatizing attitudes (Williams & Pow, 2007; Yoshioka, Reavley, MacKinnon, & Jorm, 2014), there is also a need to explore gender measurement invariance.

With an increased political interest in school-based, mental health education aiming to reduce stigma (Department of Health and Education, 2017), it is important to assess self-reported behavior change as well as hypothetical actions and intended behaviors. The aim of the current study is therefore to evaluate the quality of the RIBS for measuring reported and

Running Title: Validation of RIBS with adolescents

intended discriminatory behaviors towards individuals experiencing mental health difficulties in an adolescent sample. Our findings can inform the future use of RIBS to monitor discriminatory behaviors in adolescent populations, and evaluate school-based mental health education aiming to reduce both negative attitudes and behaviors.

Method

Design

In this psychometric validation study we assess the internal consistency, floor and ceiling effects, construct and convergent validity, content validity, and interpretability of RIBS (Terwee et al., 2007). Internal consistency is assessed through confirmatory factor analysis (CFA) and calculations of Cronbach's alpha and McDonald's Omega. Floor and ceiling effects are reported if 15% or more of the sample receive the lowest or highest possible score (Terwee et al., 2007). Assessment of construct validity is informed by existing theory that suggests negative attitudes towards individuals experiencing mental health difficulties are predictive of lower help-seeking intentions (Clement et al., 2015; Gulliver et al., 2010). Construct validity is assessed using a latent correlation with the General Help-Seeking Questionnaire (GHSQ). Convergent validity is assessed via a latent correlation with the Mental Health Knowledge Schedule (MAKS), an alternative stigma measure that aims to assess stigma-related stereotype knowledge (Evans-Lacko et al., 2010). Content validity is assessed by exploring the readability of RIBS. Finally, interpretability is explored through sub-group analyses of gender and age group, including analysis of measurement invariance.

Sample

The current study utilized follow up data from two feasibility trials conducted in 2017-2018, as part of the [blinded for review]. The feasibility trials piloted a proposed measurement framework and informed the development and adaptation of intervention materials, and the

Running Title: Validation of RIBS with adolescents

nature and scheduling of delivery for English schools, ahead of an efficacy trial (currently underway). Ethical approval was granted by [blinded for review]. Given that this study was conducted using data from two school-based trials, the sample was a convenience sample and therefore only focused on early-adolescents (year 7-8, age 11-13) and mid-adolescents (year 9-10, age 13-15).

A total of $N = 1,823$ participants were recruited from ten secondary schools in South East England. Prior to data collection, a parental opt-out consent procedure led to a total of $N = 35$ pupils being opted out of the study. Three schools were unable to facilitate the follow up survey. Pupils from the remaining seven schools completed the secure online survey in teacher facilitated sessions between June and July 2018. Seven pupils were removed that had not provided assent leaving a sample of $N = 1,132$ pupils that assented to completing the survey. A further one hundred participants were removed due to incomplete data (i.e. missing for all RIBS items). Demographic information was available for 89% of those removed. Incomplete data was equally split across males ($n = 45$) and females ($n = 44$), however, more participants in year 9-10 ($n = 69$) were deleted due to incomplete data compared with year 7-8 ($n = 20$). The final sample for analyses was $N = 1,032$, 57% of the initial number of pupils recruited, of whom 42% were male ($n = 430$) and 58% were female ($n = 602$). Thirty-four percent were in early-adolescence ($n = 347$) and 66% were in mid-adolescence ($n = 685$). This sample size exceeds the minimal recommended threshold for structural equation modelling (SEM) using the robust least squares (WLSMV) estimator (> 300) (Moshagen & Musch, 2014). Missing data patterns were tested using Little's Missing Completely at Random Test (MCAR) (Little, 1988). Data were not found to be MCAR ($\chi^2 = 149.80$, $df = 99$, $p = .001$); however, the level of missing data was $<5\%$ (.6-2.1%) (Garson, 2015). Due to low levels of item level missing data, no imputation methods were utilized in the current study; the WLSMV estimator uses all available data using pairwise present.

Measures

Gender and year group was the only demographic information collected from pupils in the current study.

Reported and Intended Behavior Scale (RIBS) (Evans-Lacko et al., 2011)

RIBS was developed to assess reported and intended behavior towards individuals experiencing a mental health problem. The authors describe the term mental health problem as a “*compromise between mental illness/mental distress*” (p.6). Individuals with a mental health problem are exemplified for respondents as “*people seen by healthcare staff*”. Items 1-4 assess self-reported behaviors: living and working with, living nearby and having a close friend experiencing a mental health problem (response options: ‘yes’ ‘no’ and ‘don’t know’). In the current study, ‘yes’ was coded as one and ‘no’ and ‘don’t know’ as zero, as per an adult validation of RIBS (Yamaguchi et al., 2014). Intended behaviors (items 5-8) are assessed in terms of participants’ willingness to have future contact across the same four contexts (response options: 1 = *disagree strongly*, 2 = *disagree slightly*, 3 = *neither agree nor disagree/don’t know*, 4 = *agree slightly*, 5 = *agree strongly*). Total scores range from 4-20, with high scores indicating more positive intended behaviors. Cronbach’s alpha has been reported at .85 with an adult sample (Evans-Lacko et al., 2011) and .86 with an adolescent sample aged 11-13 (Chisholm et al., 2016).

Mental Health Knowledge Schedule (MAKS) (Evans-Lacko et al., 2010)

MAKS was developed to assess stigma-related knowledge thought to relate to mental-health related attitudes and behaviors. Stigma related-knowledge (items 1-6) is measured in relation to help-seeking, support, employment, treatment, recovery, and recognition (response options: 1 = *strongly disagree* – 5 = *strongly agree*, where ‘don’t know’ is coded as 3). In the

current study a total score was calculated (6-30), with a higher score indicating better stigma-related knowledge. In order that agreement indicated a high score, item six *'Most people with mental health problems go to a healthcare professional to get help.'* was reverse coded.

Other MAKS items were developed to assess recognition of mental illnesses and were not included in the current study. The scale (items 1-6) was found to have moderate internal consistency (Cronbach's alpha = .65) with an adult sample (Evans-Lacko et al., 2010) and low internal consistency (Cronbach's alpha = .24) in an adolescent sample (Chisholm et al., 2016).

General Help-Seeking Questionnaire (GHSQ) (Wilson, Deane, Ciarrochi, & Rickwood, 2005)

The GHSQ consists of 10 items that assess the intention to seek help from informal sources (friend, parent and non-parent family), teachers, professional sources (mental health professional, telephone mental health helpline, doctor/GP), someone else not listed, or to not seek help at all. It was designed to be adapted to fit different samples and research contexts however, items are commonly presented in the following format: *'If you have [problem type], how likely are you to talk to a [help source] about it?'*. In the current study, pupils were presented with the following: *'please circle the number that shows how likely it is that you would seek help from each of these people if you were experiencing difficulties with your thoughts, feelings and behaviors during the next 4 weeks?'*. Participants responded to each source of help using a seven-point Likert scale (1 = *extremely unlikely* – 7 = *extremely likely*). A 'partner or significant other' item was removed from the current study as it was not deemed appropriate for all ages. Scores on items referring to specific help sources (items 1-8) were averaged to provide a mean likelihood of help-seeking. Due to the optional nature of the 'someone else not listed' item, this was not included in the calculation. When asked about intended help-seeking for personal-emotional problems, items showed moderate to good

Running Title: Validation of RIBS with adolescents

internal consistency (Cronbach's $\alpha = .70$) in a sample of 218 high school students (Wilson et al., 2005).

Data Analysis

Data handling was conducted in SPSS Version 23 and analyses in Mplus version 8.1. Models were assessed using SEM, with WLSMV estimator due to categorical item responses (Li, 2016). Controlling for less than 50 clusters when conducting two-level CFA (Type = Complex command in Mplus) can cause inaccuracy of group level parameters and standard errors (Hox, Maas, & Brinkhuis, 2010). Thus, due to the low number of schools ($N = 7$), all analyses were conducted without accounting for clustering. The criteria used for assessing good model fit was an RMSEA value of $<.06$ and CFI and TLI values $>.95$ (Hu & Bentler, 1999). Factor structure was examined by conducting CFA for two-latent factors, reported and intended behaviors, previously confirmed with university student and adult samples (Pingani et al., 2016; Yamaguchi et al., 2014). Cronbach's α and McDonald's ω were calculated to assess internal consistency. McDonald's ω supplemented α as it ensures better estimates in the absence of tau-equivalence, normality, and for data comprising ordinal responses (Trizano-Hermosilla & Alvarado, 2016). Construct validity was assessed by correlating the latent variable(s) with the mean likelihood of help-seeking calculated from the GHSQ. Convergent validity was assessed by correlating the latent variable(s) with MAKS.

Differences in responses to reported and intended behavior items were explored using group measurement invariance. Specifically, differential item functioning (DIF) comparing gender and age group was examined in a multiple indicator multiple cause (MIMIC) model. Gender and age group were added as covariates of the latent factor(s) and modification indices (MI) were used to identify direct and indirect effects. DIF between groups is identified when direct effects are observed. The MIMIC model can identify differences in thresholds and factor

means. Multi-group CFA was also conducted to explore group measurement invariance using factor loadings and residual variances. Baseline models were examined separately for each group (males, females, early and mid-adolescence) to confirm the two-factor structure. Configural and scalar invariance was assessed by freely estimating factor loadings and thresholds, and then by fixing them across groups. The scalar invariance model was then compared to the configural invariance model using the Mplus command DIFFTEST, in which a non-significant chi-square value indicates full measurement invariance. Partial invariance was explored by relaxing the parameters suggested in the MI output, and conducting a DIFFTEST with the previous model.

Readability

Readability estimates were calculated using four well established indices previously applied to other adolescent measures (Patalay, Hayes, & Wolpert, 2018): the Dale-Chall Readability Formula (DC) (Chall & Dale, 1995; Dale & Chall, 1948), the Flesch-Kincaid Reading Grade (FK) (Kincaid, Fishburne, Rogers, & Chissom, 1975), the Gunning Fog Index (GFI) (Gunning, n.d.) and the Coleman Liau Index (Coleman & Liau, 1975). Each has a unique focus, estimating readability by incorporating different assessments. For example, the formula for DC incorporates the proportion of difficult words, FK the average number of syllables per word, GFI the number of words made up of three or more syllables, and CLI the average number of letters per word. In all cases, readability was calculated as a US grade-level and converted to chronological age by adding six to the grade-level score.

Results

Factor Structure

Figure 1 presents the model with two factors ‘reported behavior’ and ‘intended behavior’.

Good model fit is evident ($N = 1,032$, $\chi^2 = 78.37$; $df = 19$; $p < .001$; RMSEA [90% CI] = .06 [.04 - .07], CFI = 1.00, TLI = 1.00). Factor loadings are significant in all cases, with generally higher loadings for intended behavior items compared with items loading onto the reported behavior factor. Item three ‘*Do you currently have, or have you ever had, a neighbour with a mental health problem?*’ demonstrated a comparatively weaker estimate compared with other reported behavior items.

[Insert Figure 1 here]

Floor/Ceiling Effects

Floor and Ceiling effects were only assessed for the intended behavior scale as the items loading onto the reported behavior scale used a ‘yes’ ‘no/don’t know’ response format, and are intended to be reported as prevalence. No floor effects were found for the intended behavior scale; however, 23% received the highest score of 20, indicative of a ceiling effect. A high score indicates more positive intended behaviors towards individuals experiencing mental health difficulties, with over one fifth of the sample strongly agreeing that they were willing to have contact across the four different contexts. At an item level, ceiling effects were found with >15% of participants selecting ‘*strongly agree*’ to all items. Table 1 shows the item response distribution.

[Insert Table 1 here]

Reliability and Validity

A high level of internal consistency was found for the intended behavior scale (Cronbach's $\alpha = .94$; McDonald's $\omega = .94$). Table 2 presents descriptive statistics, bivariate and latent correlations and measures of internal consistency for the intended behavior scale, MAKS and the GHSQ. Results reveal significant correlations between intended behavior and MAKS and between MAKS and GHSQ. More positive intended behaviors towards individuals experiencing mental health difficulties predicted greater stigma-related knowledge on MAKS; this association was moderate in magnitude ($r > .30$) (Cohen, 1988). Furthermore, the higher the score on MAKS, the higher the average help-seeking intentions on the GHSQ. However, the effect was very small ($r < .10$). No significant correlations were found between the intended behavior scale and the average GHSQ score.

[Insert Table 2 here]

Readability

Table 3 presents the four readability scores by US grade-level, the average of the four indices, and the reading age. The introductory text shows a considerably higher average reading age compared with the instructions for completion and individual items. A reading age of 19.55 indicates that the introductory text can be read by adults. Item eight '*In the future, I would be willing to continue a relationship with a friend who developed a mental health problem*' was also shown to have a higher average reading age compared with all other items, caused by sentence length and the use of more complex words. Both the reported and intended behavior sub-scales (and, as a result, the total RIBS) had an average reading age of approximately 14 years. This is considerably lower than the adult group (aged 25-45) for which the scale was originally developed and validated (Evans-Lacko et al., 2011).

[Insert Table 3 here]

Differential Item Functioning

A MIMIC model that included gender and age group as covariates in the CFA model showed acceptable model fit ($N = 1,032$, $\chi^2 = 121.77$; $df = 31$; $p < .001$; RMSEA [90% CI] = .05 [.04 - .06], CFI = 1.00, TLI = .99). No direct effects were observed for gender or age group on the reported behavior latent factor; however, there was a significant effect of gender on the intended behavior latent factor, such that females showed significantly more positive intended behaviors than males. There was also an effect of age group on the intended behavior latent factor, with mid-adolescents showing significantly lower scores than early-adolescents. Model MIs showed that adding the direct effect of age group on item one and the direct effect of gender on item four would improve the model, however fit remained consistent ($N = 1,032$, $\chi^2 = 96.28$; $df = 29$; $p < .001$; RMSEA [90% CI] = .05 [.04 - .06], CFI = 1.00, TLI = .99). See Figure 1 for direct and indirect effects.

Measurement Invariance

Table 4 presents the model fit indices for baseline models, configural and scalar measurement invariance. The female and early-adolescent baseline models showed good model fit. However, the RMSEA was less satisfactory for males and mid-adolescents. Model fit indices indicated full configural measurement invariance for gender and age group. Full scalar measurement invariance was not found, however, acceptable model fit indices indicated the possibility of partial scalar measurement invariance. MIs indicated that relaxing the factor loading and threshold of item eight '*In the future, I would be willing to continue a relationship with a friend who developed a mental health problem*' in the gender model would improve fit. After relaxing these parameters, partial scalar invariance was achieved. Similarly, relaxing the factor loadings and thresholds for items five '*In the future, I would be*

Running Title: Validation of RIBS with adolescents

willing to live with someone with a mental health problem' and seven *'In the future, I would be willing to live nearby to someone with a mental health problem*' achieved partial scalar invariance in the age group model.

[Insert Table 4 here]

Discussion

This psychometric validation study assessed the quality of RIBS for measuring self-reported and intended discriminatory behavior in an adolescent sample. We conclude that RIBS is a valid measure of self-reported and intended discriminatory behavior for an adolescent sample. However, caution should be taken when using the scale with adolescents under the age of 14 with regards to its readability, and when exploring mean differences for gender and age group.

The full scale (items 1-8) was found to have an average reading age of approximately 14 years old, equivalent to US grade eight. All readability scores were above the youngest participants (11 years old), with only two items in the 11-year old range (11.12 and 11.21) and two in the 12-year old range (both 12.72). Other items range from 13.01-16.91 years, exceeding the recommended reading age for adult measures (12 years old) (Terwee et al., 2007). Of particular concern was the introductory text which exceeded an average reading age of 19. Given the variation in reading ability in any given classroom, and that a proportion of young people will fall below the reading age for their grade, caution must be taken in terms of content validity and respondent burden for adolescents. One possible solution is to adapt the introductory text, simplifying language and content (e.g. *'mental'* *'problems'* and *'healthcare'*). Given the reference to *'people who have mental health problems'* in all items, simplifying the introductory text could improve the readability of the whole scale.

Furthermore, the reference to service use here undoubtedly influences the perceived level of

mental distress. Evidence suggests that stigma is a “*matter of degree*” (Link et al., 2004, p.513) influenced by labels and perceived severity (Jorm & Griffiths, 2008). Mental health related terms such as ‘mental health problem’, ‘mental distress’ and ‘mental illness’ can produce different stigma responses due to common associations and misconceptions. For example, schizophrenia is more likely to be associated with the term ‘mental illness’ than depression (Evans-Lacko et al., 2010). Leighton (2009) reported inconsistencies in young people’s understanding of mental health related terms, in part due to the lack of professional clarity, leading to increased subjectivity in their meaning.

A minority of the sample reported living and working with, living nearby, and having a close friend experiencing a mental health problem. Early-adolescent participants were less likely than mid-adolescents to report having lived with someone experiencing a mental health problem. Furthermore, males were less likely than females to report having a close friend experiencing a mental health problem. In contrast to reported behaviors, ceiling effects were found at a scale and item level for intended behaviors, indicative of positive attitudes among participants in terms of contact willingness across the four different contexts.

These ceiling effects support findings from previous adult literature (Evans-Lacko et al., 2011). For example, more normally distributed responses were found for items describing closer social contact (e.g. living with someone) compared with items describing more distant relationships (e.g. living nearby). Responses were skewed towards agreement for item eight despite high levels of social contact, indicating that the majority of participants would ‘*continue a relationship with a friend who developed a mental health problem*’. As suggested by Evans-Lacko et al. (2011) responses to item eight may be mediated by higher reports of having a friend experiencing a mental health problem in item four. Despite showing more normally distributed results overall, similar patterns of responses were previously found in Japanese and Italian adult validation studies (Pingani et al., 2016; Yamaguchi et al., 2014).

Running Title: Validation of RIBS with adolescents

Overall, increased levels of agreement for intended behaviors involving less social contact, and lower levels of reported behaviors, indicate more social desirability for items describing hypothetically more distant relationships. The difference in responses between the reported and intended behavior items is also further support that hypothetical behavior may not always translate into actual behavior (Eisenberg et al., 2012; Thornicroft et al., 2007).

Results from the two-factor CFA showed good model fit, confirming the previously identified factors (Pingani et al., 2016; Yamaguchi et al., 2014). Chi-square is highly sensitive to sample size, in which a large sample size is more likely to produce a significant result for a small effect (Vandenberg, 2006). Therefore, despite a significant chi-square value, model fit was perceived to be good based on the remaining indices and high factor loadings. Furthermore, a high level of internal consistency was found for the intended behavior scale (.94). This value was higher than previous assessments in adult (.85) (Evans-Lacko et al., 2011) and adolescent research (.86) (Chisholm et al., 2016).

A significant but moderate positive correlation was found between the intended behavior scale and MAKS, supporting the relationship between stigma-related knowledge and attitudes (Thornicroft et al., 2007). However, contrary to our predictions, there was no significant correlation between intended behavior and help-seeking intentions measured by the GHSQ. Previous literature suggests that stigmatizing attitudes reduce the likelihood of seeking help (Gulliver et al., 2010), in particular, internalized stigma and stigma relating to treatment (Clement et al., 2015). Given that RIBS is a measure of self-reported and intended discriminatory behavior, revealing attitudes towards others, it may be more likely to relate to *providing* support as opposed to seeking it.

Gender and age group had significant effects on the intended behavior latent factor, with females and early-adolescents reporting more positive intended behaviors overall. The former

finding supports previous literature that suggests that males report more stigmatizing attitudes than females (Williams & Pow, 2007; Yoshioka et al., 2014). The more negative intended behaviors reported by mid-adolescents add to complex age trends found in previous stigma research. Jorm and Wright (2008) found that social distance reduced with age, however, their study also revealed that increased age predicted higher levels of perceived stigma and the belief that people experiencing mental health difficulties are dangerous and unpredictable, which could in turn influence intended behaviors. In the current study, it is possible that more positive intended behaviors reported by the early-adolescents could be a result of greater social desirability or, given the reading age, reduced validity due to a lack of comprehension.

Multi-group CFA revealed full configural measurement invariance for gender and age group, indicating the same item-factor structure i.e. 4 items loaded onto each latent factor, across groups. However, only partial scalar measurement invariance was achieved for gender and age group. Non-invariance across groups was found for a small number of item factor loadings and thresholds, however, these did not relate to overall increased levels of intended behaviors in females and early-adolescents. Given that full scalar invariance was not achieved, it could be recommended that direct comparisons of group means might not be meaningful. In particular, it has been shown that invariance at the scalar level has large effects on the accuracy of mean-level analyses (Steinmetz, 2013). However, many researchers are managing group non-invariance by relaxing constraints on parameters identified in the modification indices. More research is needed to fully understand the statistical consequences of accepting partial invariance (Putnick & Bornstein, 2016).

However, the utilization of measurement invariance analyses in the current study contributes to the growing body of literature that recognizes the importance of assessing the uniformity of the psychometric properties of a construct across groups before conducting comparisons of means.

Running Title: Validation of RIBS with adolescents

Though the sample size in the current study is large, it is worth noting that limited demographic information was available for the participants. We therefore cannot make any statements about representativeness in terms of ethnicity and socio-economic factors compared to the English population. Future research should aim to collect data on these variables to study differential item functioning. Furthermore, the convenience sample was limited to early and mid-adolescents, therefore future research should consider expanding the sample to include adolescents aged 16 years and older and compare item functioning. Follow up data from two intervention feasibility trials were utilized in this study, however, RIBS was not the proposed primary outcome for interventions.

Readability formulae were used in the current study to provide an additional assessment of the appropriateness of RIBS for adolescents. Despite the use of multiple indices, the authors acknowledge that when taken alone, readability assessments can lack reliability, particularly when applied to short extracts of text (Oakland & Lane, 2004). The high readability scores in the current study should therefore be taken with some caution, particularly given the results from other assessments of psychometric quality. Future research should use qualitative approaches such as focus groups and cognitive interviewing to explore adolescents' experiences and interpretation of RIBS, to reveal any issues with readability and inform future revisions to the scale. For example, accounting for developmental stage by changing "work" to "school" in items 2 and 6.

Convergent validity was assessed by correlating the intended behavior scale with MAKS, an alternative stigma measure aiming to assess stigma-related stereotype knowledge. MAKS was only found to have moderate levels of internal consistency in the current sample, suggesting that it may not function well as a unidimensional scale in this age group. Future research should conduct a full assessment of the psychometric properties of MAKS with an adolescent sample. It was also not possible in the current study to assess all possible criteria

Running Title: Validation of RIBS with adolescents

for assessing the psychometric quality of RIBS. For example, data from only one time point was available and therefore no assessment of test-retest reliability was conducted.

The current study concludes that RIBS is a valid measure for adolescents aged 14+ when the two-factor structure of reported and intended discriminatory behaviors is utilized. However, the introductory text is not appropriate for this age group due to its high reading age and would benefit from being simplified. Researchers should anticipate discrepancies between the reported and intended behavior response distributions, and test for possible ceiling effects in their sample. Gender and age group measurement invariance should also be assessed prior to mean difference testing on the intended behavior scale.

References

- Chall, J. S. (Jeanne S., & Dale, E. (1995). *Readability revisited : the new Dale-Chall readability formula*. Retrieved from <https://www.worldcat.org/title/readability-revisited-the-new-dale-chall-readability-formula/oclc/32347586>
- Chisholm, K., Patterson, P., Torgerson, C., Turner, E., Jenkinson, D., & Birchwood, M. (2016). Impact of contact on adolescents' mental health literacy and stigma: the SchoolSpace cluster randomised controlled trial. *BMJ Open*, *6*(2), e009435. <https://doi.org/10.1136/bmjopen-2015-009435>
- Ciarrochi, J., Morin, A. J. S., Sahdra, B. K., Litalien, D., & Parker, P. D. (2017). A longitudinal person-centered perspective on youth social support: relations with psychological wellbeing. *Developmental Psychology*, *53*(6), 1154–1169. <https://doi.org/10.1037/dev0000315>
- Clement, S., Schauman, O., Graham, T., Maggioni, F., Evans-Lacko, S., Bezborodovs, N., ... Thornicroft, G. (2015). What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. *Psychological Medicine*, *45*(1), 11–27. Retrieved from <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=psyc12&AN=2014-54680-002>
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*.
- Coleman, M., & Liau, T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, *60*(2), 283–284. <https://doi.org/10.1037/h0076540>

Running Title: Validation of RIBS with adolescents

Corrigan, P. (2000). Mental health stigma as social attribution: implications for research methods and attitude change. *American Psychological Association*, 7(1), 1–20.

<https://doi.org/10.1093/clipsy.7.1.48>

Corrigan, P. W., Morris, S. B., Michaels, P. J., Rafacz, J. D., & Rusch, N. (2012).

Challenging the public stigma of mental illness: A meta-analysis of outcome studies.

Psychiatric Services, 63(10), 963–973.

Crush, E., Arseneault, L., Moffitt, T. E., Danese, A., Caspi, A., Jaffee, S. R., ... Fisher, H. L.

(2018). Protective factors for psychotic experiences amongst adolescents exposed to multiple forms of victimization. *Journal of Psychiatric Research*, 104, 32–38.

<https://doi.org/10.1016/j.jpsychires.2018.06.011>

Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research*

Bulletin, 27(21), 37–54. <https://doi.org/10.2753/JEI0021-3624440403>

DeLuca, J. S. (2019). Conceptualizing adolescent mental illness stigma: youth stigma

development and stigma reduction programs. *Adolescent Research Review*.

<https://doi.org/10.1007/s40894-018-0106-3>

Department of Health and Education. (2017). *Transforming children and young people's*

mental health provision: a green paper. <https://doi.org/979-1-5286-0061-3>

Eisenberg, D., Speer, N., & Hunt, J. B. (2012). Attitudes and beliefs about treatment among

college students with untreated mental health problems. *Psychiatric Services*, 63(7), 6–8.

Evans-lacko, S., Little, K., Meltzer, H., Rose, D., Rhydderch, D., Henderson, C., &

Thornicroft, G. (2010). Development and psychometric properties of the Mental Health

Knowledge Schedule. *The Canadian Journal of Psychiatry*, 55(7), 440–448.

Running Title: Validation of RIBS with adolescents

Evans-Lacko, S., Rose, D., Little, K., Flach, C., Rhydderch, D., Henderson, C., &

Thornicroft, G. (2011). Development and psychometric properties of the Reported and Intended Behaviour Scale (RIBS): a stigma-related behaviour measure. *Epidemiology and Psychiatric Sciences*, 20(03), 263–271.

<https://doi.org/10.1017/S2045796011000308>

Fox, A. B., Earnshaw, V. A., Taverna, E. C., & Vogt, D. (2017). Conceptualizing and measuring mental illness stigma: the mental illness stigma framework and critical review of measures. *Stigma and Health*. <https://doi.org/10.1037/sah0000104>

Garson, G. D. (2015). Missing values analysis and data imputation. *Statistical Associates Publishers*, 1–26.

Griffiths, K. M., Christensen, H., & Jorm, A. F. (2008). Predictors of depression stigma. *BMC Psychiatry*, 8, 1–12. <https://doi.org/10.1186/1471-244X-8-25>

Gulliver, A., Griffiths, K. M., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC Psychiatry*, 10.

Gunning, F. (n.d.). The technique of clear writing. Retrieved from

<http://www.amazon.com/technique-clear-writing-Robert-Gunning/dp/B0006ASTRK>

Hox, J. J., Maas, C. J. M., & Brinkhuis, M. J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, 64(2), 157–170. <https://doi.org/10.1111/j.1467-9574.2009.00445.x>

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>

- Jorm, A. F., & Griffiths, K. M. (2008). The public's stigmatizing attitudes towards people with mental disorders: how important are biomedical conceptualizations? *Acta Psychiatrica Scandinavica*, *118*(4), 315–321. <https://doi.org/10.1111/j.1600-0447.2008.01251.x>
- Jorm, A. F., & Wright, A. (2008). Influences on young people's stigmatising attitudes towards peers with mental disorders: national survey of young Australians and their parents. *The British Journal of Psychiatry*, *192*(2), 144–149.
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. <https://doi.org/10.21236/ADA006655>
- Leighton, S. (2009). Adolescents' understanding of mental health problems: conceptual confusion. *Journal of Public Mental Health*, *8*(2), 4–14.
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
- Link, B. G., Yang, L. H., Phelan, J. C., & Collins, P. Y. (2004). Measuring mental illness stigma. *Schizophrenia Bulletin*, *30*(3), 511–541. <https://doi.org/10.1093/oxfordjournals.schbul.a007098>
- Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of The American Statistical Association*, *83*(404), 1198–1202.
- Moses, T. (2010). Being treated differently: stigma experiences with family, peers, and school staff among adolescents with mental health disorders. *Social Science & Medicine*, *70*(7), 985–993.

Running Title: Validation of RIBS with adolescents

Moshagen, M., & Musch, J. (2014). Sample size requirements of the robust weighted least squares estimator. *Methodology*, *10*(2), 60–70. <https://doi.org/10.1027/1614-2241/a000068>

Patalay, P., Hayes, D., & Wolpert, M. (2018). Assessing the readability of the self-reported Strengths and Difficulties Questionnaire. *BJPsych Open*, *4*(02), 55–57. <https://doi.org/10.1192/bjo.2017.13>

Pingani, L., Evans-Lacko, S., Luciano, M., Del Vecchio, V., Ferrari, S., Sampogna, G., ... Fiorillo, A. (2016). Psychometric validation of the Italian version of the Reported and Intended Behaviour Scale (RIBS). *Epidemiology and Psychiatric Sciences*, *25*(5), 485–492. <https://doi.org/10.1017/S2045796015000633>

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>.Measurement

Steinmetz, H. (2013). Analyzing observed composite differences across groups: is partial measurement invariance enough? *Methodology*, *9*(1), 1–12. <https://doi.org/10.1027/1614-2241/a000049>

Stuart, H., Arboleda-Florez, J., & Sartorius, N. (2012). *Paradigms lost: fighting stigma and the lessons learned*. Oxford University Press.

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., ... de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>

Running Title: Validation of RIBS with adolescents

Thornicroft, G. (2008). Stigma and discrimination limit access to mental health care.

Epidemiologia e Psichiatria Sociale, 17(1), 14–19.

<https://doi.org/10.1017/S1121189X00002621>

Thornicroft, G., Rose, D., & Kassam, A. (2007). Stigma: ignorance, prejudice or

discrimination? *British Journal of Psychiatry*, 190, 192–193. <https://doi.org/10.1192/bjp.bp.106.02525791> AUTHOR

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha

reliability in realistic conditions: congeneric and asymmetrical measurements. *Frontiers*

in Psychology, 7, 1–8. <https://doi.org/10.3389/fpsyg.2016.00769>

Van Droogenbroeck, F., Spruyt, B., & Keppens, G. (2018). Gender differences in mental

health problems among adolescents and the role of social support: results from the

Belgian health interview surveys 2008 and 2013. *BMC Psychiatry*, 18(1), 1–9.

<https://doi.org/10.1186/s12888-018-1591-4>

Vandenberg, R. J. (2006). Statistical and methodological myths and urban legends: where,

pray tell, did they get this idea? *Organizational Research Methods*, 9(2), 194–201.

<https://doi.org/10.1177/1094428105285506>

Wei, Y., Hayden, J. A., Kutcher, S., Zygmunt, A., & McGrath, P. (2013). The effectiveness

of school mental health literacy programs to address knowledge, attitudes and help

seeking among youth. *Early Intervention in Psychiatry*, (7), 109–121.

<https://doi.org/10.1111/eip.12010>

Wei, Y., McGrath, P., Hayden, J., & Kutcher, S. (2017). The quality of mental health literacy

measurement tools evaluating the stigma of mental illness: a systematic review.

Epidemiology and Psychiatric Sciences, 1–30.

Running Title: Validation of RIBS with adolescents

<https://doi.org/doi:10.1017/S2045796017000178>

Williams, B., & Pow, J. (2007). Gender differences and mental health: an exploratory study of knowledge and attitudes to mental health among Scottish teenagers. *Child and Adolescent Mental Health*, 12(1), 8–12. <https://doi.org/10.1111/j.1475-3588.2006.00413.x>

Wilson, C. J., Deane, F. P., Ciarrochi, J., & Rickwood, D. (2005). Measuring Help-Seeking Intentions: Properties of the General Help-Seeking Questionnaire. *Canadian Journal of Counselling*, 39(1), 15–28. [https://doi.org/Retrived from http://cjc-rcc.ucalgary.ca/cjc/index.php/rcc/article/view/265](https://doi.org/Retrived%20from%20http://cjc-rcc.ucalgary.ca/cjc/index.php/rcc/article/view/265)

Yamaguchi, S., Koike, S., Watanabe, K. I., & Ando, S. (2014). Development of a Japanese version of the Reported and Intended Behaviour Scale: reliability and validity. *Psychiatry and Clinical Neurosciences*, 68(6), 448–455. <https://doi.org/10.1111/pcn.12151>

Yoshioka, K., Reavley, N. J., MacKinnon, A. J., & Jorm, A. F. (2014). Stigmatising attitudes towards people with mental disorders: results from a survey of Japanese high school students. *Psychiatry Research*, 215(1), 229–236.

Running Title: Validation of RIBS with adolescents

Figure 1

The Two-Factor Structure for RIBS and MIMIC Models with Gender and Age Group as Covariates

[see Figure 1.pdf]

*Note: b) is without direct effects and c) includes direct effects. Models include standardized path coefficients (standard error), factor loadings (standard error) and correlations between factors. Note that ** indicates parameter estimate with $p \leq .001$ and * $p < .010$.*

Table 1

Response Distribution for the Reported and Intended Behavior Scales

Reported Behavior Scale (% (count) responding yes)	Male	Female	Early-adolescents	Mid-adolescents	Total
1) Are you currently living with, or have you ever lived with, someone with a mental health problem?	16(68)	16(98)	11(39)	19(127)	16(166)
2) Are you currently working with, or have you ever worked with, someone with a mental health problem?	18(76)	16(96)	17(58)	17(114)	17(172)
3) Do you currently have, or have you ever had, a neighbour with a mental health problem?	14(61)	12(72)	14(46)	13(87)	13(133)
4) Do you currently have, or have you ever had, a close friend with a mental health problem?	22(92)	32(188)	29(99)	27(181)	28(280)
Intended Behavior Scale (% (count) total sample)	1	2	3	4	5
5) In the future, I would be willing to live with someone with a mental health problem.	05(51)	07(73)	41(417)	22(227)	25(255)
6) In the future, I would be willing to work with someone with a mental health problem.	03(28)	03(33)	30(308)	27(281)	37(375)
7) In the future, I would be willing to live nearby to someone with a mental health problem.	03(26)	02(23)	29(289)	27(268)	40(404)
8) In the future, I would be willing to continue a relationship with a friend who developed a mental health problem.	02(22)	03(30)	30(305)	22(224)	43(435)

Note: 1 = disagree strongly, 2 = disagree slightly, 3 = neither agree nor disagree/don't know, 4 = agree slightly, 5 = agree strongly

Table 2

Descriptive Statistics by Gender and Age Group, Total Sample Bivariate (Latent) Correlations and Internal Reliability

Measure	Sample Size Mean(Standard Deviation)					Bivariate(Latent) Correlation Coefficients			Internal Reliability	
	Male	Female	Early- adolescents	Mid- adolescents	Total	1	2	3	A	ω
1. IBS	N = 411 14.43(3.70)	N = 587 16.25(3.46)	N = 338 16.32(3.65)	N = 660 15.08(3.62)	N = 998 15.50(3.68)	-	-	-	.94	.94
2. MAKS	N = 398 20.78(3.08)	N = 564 21.12(2.87)	N = 332 21.34(2.82)	N = 630 20.79(3.02)	N = 962 20.98(2.96)	.39**(.43**)	-	-	.53	.62
3. GHSQ_A	N = 380 3.43(1.36)	N = 537 3.31(1.17)	N = 312 3.33(1.14)	N = 605 3.38(1.30)	N = 917 3.36(1.25)	.03(.04)	.06(.06*)	-	.85	.87

Note: IBS = 4-item (items 5-8) Intended Behavior Scale, MAKS = 6-item (items 1-6) Mental Health Knowledge Schedule, GHSQ_A = Average

*Help-seeking ((total items 1-8)/8), ** $p < .001$, * $p \leq .05$.*

Table 3

Average Readability Estimates (US Grade and Years) for the Introductory Text, Instructions, Full Scale, Sub-Scales and Items of RIBS

	DC	FK	GFI	CLI	Grade	Age
Introductory Text						
The following questions ask about your experiences and views in relation to people who have mental health problems (for example, people seen by healthcare staff).	9.30	13.98	16.40	14.53	13.55	19.55
Instructions: For each of questions 1–4, please respond by ticking one box only.	5.46	5.09	5.60	4.34	5.12	11.12
1) Are you currently living with, or have you ever lived with, someone with a mental health problem?	7.27	7.68	9.15	9.43	8.38	14.38
2) Are you currently working with, or have you ever worked with, someone with a mental health problem?	7.27	7.68	9.15	10.13	8.56	14.56
3) Do you currently have, or have you ever had, a neighbour with a mental health problem?	7.39	6.93	8.90	7.34	7.64	13.64
4) Do you currently have, or have you ever had, a close friend with a mental health problem?	7.27	6.26	9.15	6.67	7.34	13.34
Reported Behaviors Sub-Scale Total	7.30	7.11	9.09	8.41	7.97	13.97
Instructions: For each of the questions 5–8, please respond by ticking one box only.	5.43	5.25	6.00	4.18	5.21	11.21
5) In the future, I would be willing to live with someone with a mental health problem.	7.39	6.11	6.40	6.97	6.72	12.72
6) In the future, I would be willing to work with someone with a mental health problem.	7.39	6.11	6.40	6.97	6.72	12.72
7) In the future, I would be willing to live nearby to someone with a mental health problem.	7.27	6.97	6.80	7.01	7.01	13.01
8) In the future, I would be willing to continue a relationship with a friend who developed a mental health problem.	9.37	10.50	14.00	9.77	10.91	16.91
Intended Behaviors Sub-Scale Total	7.92	7.54	8.64	7.79	7.97	13.97
Introductory Text + Instructions	7.16	8.66	10.16	9.01	8.75	14.75
Total RIBS (items 1-8)	7.61	7.32	8.86	8.09	7.97	13.97

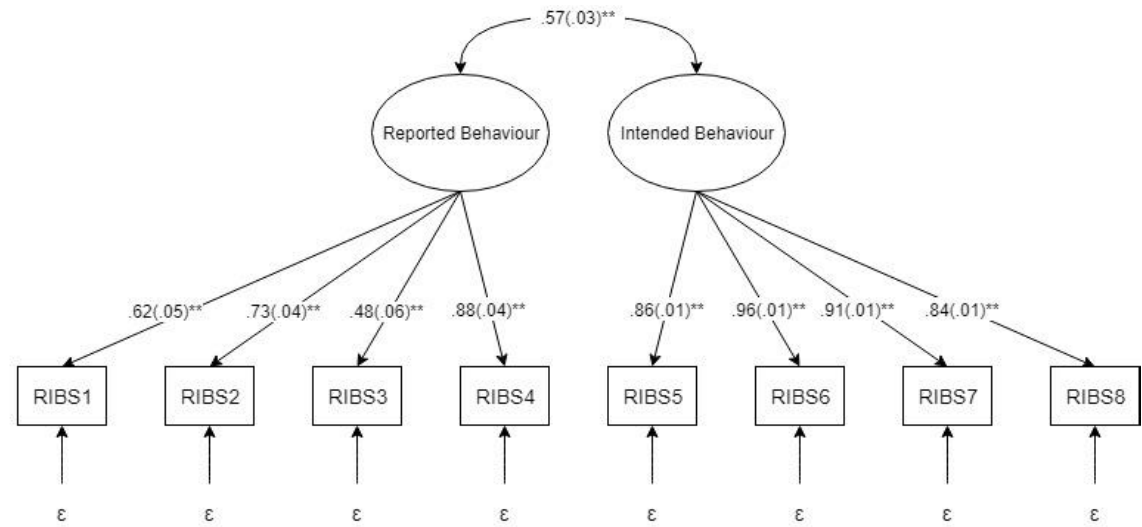
Table 4

Multi-Group CFA for Gender and Age Group

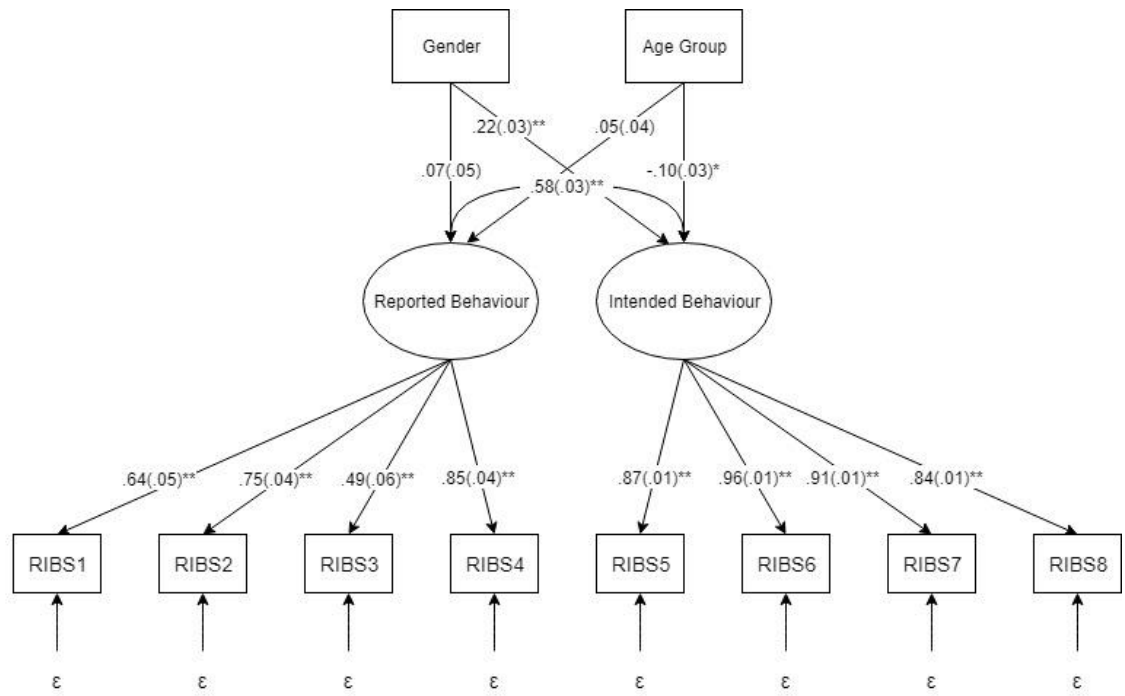
Models	χ^2	df	RMSEA [90% CI]	CFI	TLI	χ^2 diff (df)
Gender						
Baseline Males	70.27**	19	.08 [.06 - .10]	.99	.99	n/a
Baseline Females	30.09*	19	.03 [.00 - .05]	1.00	1.00	n/a
Configural	98.22**	38	.06 [.04 - .07]	1.00	1.00	n/a
Scalar	124.49**	54	.05 [.04 - .06]	1.00	1.00	30.62(16)*
Scalar – RIBS8	110.48**	50	.05 [.04 - .06]	1.00	1.00	17.05(12)
Age group						
Baseline Early-adolescents	30.10*	19	.04 [.00 - .07]	1.00	1.00	n/a
Baseline Mid-adolescents	74.97**	19	.07 [.05 - .08]	1.00	.99	n/a
Configural	96.13**	38	.05 [.04 - .07]	1.00	1.00	n/a
Scalar	138.91**	54	.06 [.04 - .07]	1.00	1.00	46.13(16)**
Scalar – RIBS5, RIBS7	98.54**	46	.05 [.03 - .06]	1.00	1.00	9.04(8)

*Note: CFA = confirmatory factor analysis; RMSEA = Root Mean Square Error of Estimation; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index, RIBS5, RIBS7 and RIBS8 refer to the items for which factor loadings and thresholds were released to achieve partial scalar invariance, ** $p < .001$, * $p \leq .05$.*

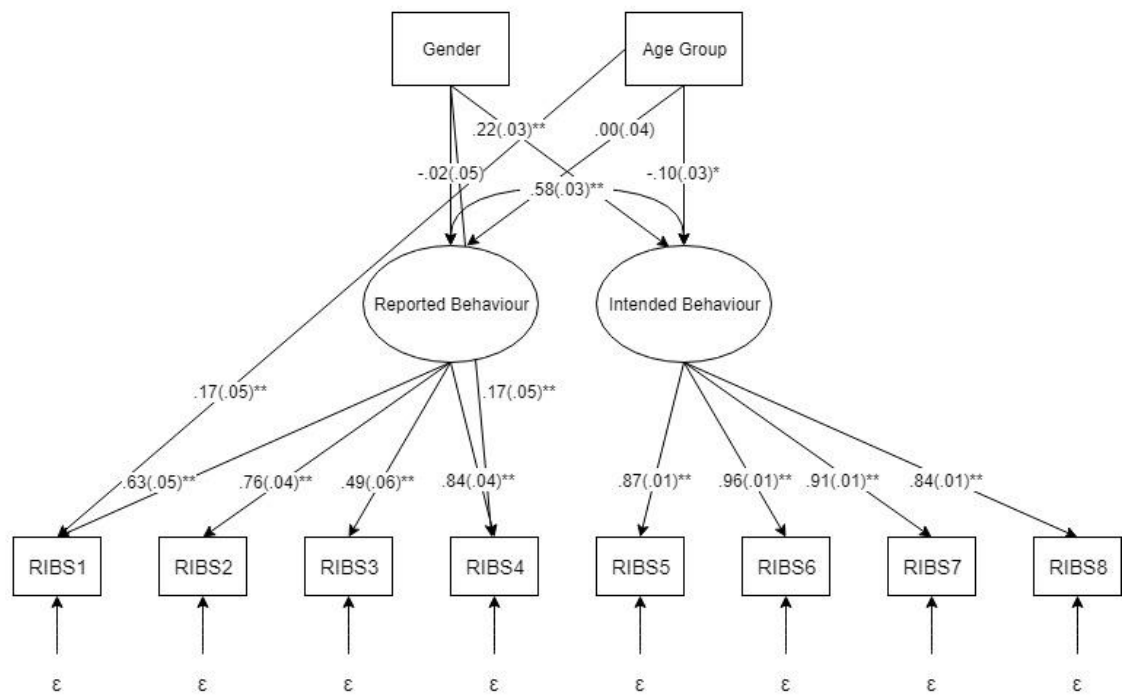
Figure



a)



b)



c)