Initial Assessment and Referral Decision Support Tool (IAR-DST)

Validity Study: Final Report

INSIDE OUT
Institute for Eating Disorders

© InsideOut Institute, 2025

Acknowledgement of Country

InsideOut Institute acknowledges Aboriginal and Torres Strait Islander peoples as the Traditional Custodians of the lands on which we live and work - we acknowledge their continuing connection to, and their care of the lands, water, sea and sky since time immemorial. We recognise Aboriginal and Torres Strait Islander people's culture is the oldest living culture in human history, and we pay our respects to Elders past, present, and future leaders. This always was and always will be Aboriginal land.

Recognition of Lived Experience

We recognise the individual and collective experience of those with lived and/or living experiences of mental ill-health and recovery, including their families, supporters and carers. We acknowledge their courageous contribution by sharing their unique and diverse expertise in the hope of improving better outcomes for all. We value and are committed to their partnership, placing lived experience expertise at the heart of research as we develop and shape our work.

Funding

The project was funded by the Australian Government Department of Health, Disability and Ageing, 2024-2025.

Disclaimer

InsideOut Institute has prepared this report for the benefit of the Australian Government Department of Health, Disability and Ageing. This report should not be used or relied upon for any purpose other than as an expression of the conclusions and recommendations of InsideOut in relation to the scope of the report. This report has been prepared by InsideOut based on information provided by the Department of Health, Disability and Ageing and by other persons and organisations, as well as through its own research. We relied on that information and have not independently audited or verified that information. Maguire, S. & Marks, P. designed the study, Nassar, N. & Schneuer, F. conducted data analyses, Dann, K.M, Vatter, S. & Maguire, S. conducted the research and wrote the report.

Recommended Citation

InsideOut Institute (2025) Initial Assessment and Referral Decision Support Tool (IAR-DST) validity study: Final report. InsideOut Institute, Sydney, 2025.

Inside Out Institute is a joint venture between
The University of Sydney and Sydney Local Health District.

Contents:

Executive Summary **4**

Overview of recommendations **5**

Summary of reliability and validity findings **6**

IAR Validity project activities: Results **12**

Activity 1: Rapid desktop review of existing IAR-DST evaluations and reviews 12

Activity 2: Validity and reliability of the IAR-DST from existing data sources 17

Mapping IAR-DST Level of Care to K10+ level of distress 21

Mapping K10+ scores to IAR-DST Domain 1 symptom severity and distress scores 24

Rapid Validity Studies **27**

Rapid Study 1 27

Rapid Study 2 29

Domain 1: Symptom severity and distress 35

Domain 2: Harm 36

Domain 3: Functioning 37

Domain 4: Impact of co-existing conditions 38

Domain 5: Service use and response history 39

Domain 6: Social and environmental stressors 40

Domain 7: Family and other supports 41

Domain 8: Engagement and motivation 42

Rapid Study 3 45

References **48**

Executive Summary:

Overview of evaluation results

The **Initial Assessment and Referral Decision Support Tool (IAR-DST)** is a clinical tool that assists health professionals to recommend an appropriate level of care for those seeking mental health support.

Levels of care are based on resource intensity: Level 1, Self-Management; Level 2, Low-intensity services; Level 3, Moderate-intensity services; Level 4, High-intensity services and Level 5, Acute and Specialist Community Mental Health Services. The recommendation is based on an assessment of a person's presenting issues across eight domains, including: Symptom severity and distress, Harm, Functioning, impact of coexisting conditions, Service use and response history, Social and environmental stressors, Family and other supports and Engagement and motivation.

The **IAR-DST Validity project** is a set of pragmatic validity and reliability studies of the IAR-DST, which has been implemented across multiple settings and jurisdictions in Australia. The studies were performed within a 9-month timeframe commencing late 2024 and concluding mid-2025. This report provides the results of these studies, and recommendations on any changes that could be made to the IAR-DST to increase its validity. This independent evaluation examined and found:

A review of past evaluations of the IAR-DST provided some evidence of the simplest forms of validity including *face validity* (the degree to which the tool appears to reflect the constructs it is trying to measure) and *content validity* (the degree to which the content of the tool is an adequate reflection of the constructs it is trying to measure) of the IAR-DST. However, none evaluated the more complex aspects of validity and reliability of the tool in the settings it is being used; these tests of a tool or instrument are important because they indicate whether an instrument measures in real settings the thing that it is meant to measure, and that it does so consistently across different users and across time.

Review of online feedback on the IAR-DST from 2022-2024 revealed 64% of respondents reviewed the tool positively. Common recommendations for improvement include integration of the tool with practice software, referrals and links to services and improved user experience through adding pop-ups or hovers for guidance notes.

Analysis of the Primary Mental Health Care Minimum Dataset (PMHC-MDS) aggregated reports indicates high levels of agreement (89%) between IAR-DST recommended Level of Care and the practitioner-assigned Level of Care, however a sizeable proportion of Level 1 and 2 recommendations are being upgraded. This impedes the utility of the tool to contribute to the aims of stepped care, and referral to low-intensity interventions. Analysis of this dataset also reveals that the majority of people assessed on the IAR-DST across the country are assessed to be a Level 3, suggesting that the tool may not be contributing as it was intended to the earlier identification of people at lower levels of need and the facilitation of earlier care. However, this may also be because people are not presenting to care at an early stage.

Alignment between the IAR-DST Recommended Level of Care and the K10+, which is another widely used instrument to assess mental ill health, was poor. For those assigned the lowest level of care need by the IAR-DST (Level 1) still approximately half the sample scored in the high or very high range on the K10+, and for all other levels (2-5) the majority of the samples scored in the high or very high range on the K10+. However, this result is more likely a reflection of the limited ability of the K10+ to discriminate between level of severity of illness, than it is a reflection of the IAR-DST.

An observational study of IAR-DST encounters at a Medicare Mental Health Centre revealed a very high rate of agreement (91%) between the level assigned by intake officers and the level assigned by clinical staff, however, found no significant relationship between level of care indicated by the tool and the number of referrals given, suggesting a potential misalignment between the tool and referral options and/or practices.

A rapid validity study completed at InsideOut Institute found that across the assessment of 100 individuals, a clinician consistently rated levels of care higher than the intake officer without clinical training, suggesting the training and clinical experience of the IAR-DST user may be a significant factor affecting the reliability of the tool. The study also found only fair-to-moderate concordance between self-reported instruments measuring the severity of psychological illness and IAR-DST Domain ratings. This suggests that the accuracy of the tool could be improved if well-validated self-report measures were able to be included in the assessment process.

The majority of participants interviewed reported that the questions asked during IAR-DST assessment covered all areas of their support needs (91%) and the overall process made them feel they could get appropriate support for their needs (84%) suggesting that the structure and content of the IAR-DST are acceptable to people with a broad range of mental health concerns.

A final study **surveyed clinicians who have used the IAR-DST** to make or receive referrals from the IAR-DST, about their experiences of using the tool. The majority of clinicians who responded to the survey (n=94) were those who made referrals, and while the group was mixed, over 50% were GPs who were highly experienced in mental health. The majority of respondents had negative appraisals of the tool and its utility in their practice. They did provide useful feedback on ways the tool could be improved.

Overview of recommendations

The IAR-DST has been well embedded into the processes and practices in many primary care settings across the country. A shared language about level of care needs has emerged and is being used and shared in health setting contexts. The IAR-DST is addressing a genuine need for comprehensive needs assessment at intake for treatment for a mental health condition and a mechanism to assign based on need. Were it to be removed it would create significant disruption within primary care, and that need and gap would remain, and be exposed. There are a number of issues with the way the tool is integrated and used in some settings, its ability to differentiate sufficiently, linkages with properly mapped and accessible care options, and with uptake by certain clinician groupings, all of which can be addressed.



Incorporating patient self-reported items into the IAR-DST process following a comprehensive co-design and user-testing process within settings **will likely improve the validity and reliability of the instrument in practice** and may also improve the experience of the instrument, making it a shared care and decision-making encounter.



Creating a pathway for GP use of the tool that is compatible with, and preferably embedded in, their workflows and software could increase acceptability of the tool in GP settings. The IAR-DST could meet its objective of providing a 'shared language' for levels of care , while becoming a useful tool for upskilling novice GPs and other clinicians without overburdening highly experienced GPs.



The treatment options, referral databases and mechanisms linked to the IAR-DST very likely affect its utility, validity and reliability. The tool will only ever be as good as the treatment ecosystem it operates within. Lengthy assessments frustrate clinicians (and patients) when there is no viable referral pathway, and the current evaluation suggests that, at times, the level of care may be adjusted by users to access available care options, even if they do not align with patient need. Development of accessible treatment options matching each level of need and linkage of these pathway options to the IAR-DST encounter will improve the reliability and validity of the instrument in practice and improve perceptions of its utility.



The **breadth and domains** of the tool when assessed comprehensively in an interview seem to **reflect well the client's perceived areas of need**; overall the framework the tool adopts was received well by clients interviewed as part of this research, suggesting the tool has innate utility and worth. Future improvements, adaptations and/or integrations of the tool should take account of this and try to preserve it.



It was beyond the scope of the current study to evaluate the reliability, validity and acceptability of the IAR-DST in priority populations. **Further studies in specific populations are needed** in particular in groups belonging to different cultural communities, including Aboriginal and Torres Strait Islander peoples and those from culturally and linguistically diverse backgrounds. The IAR-DST may work very differently in these populations and may require adaptation to meet acceptability and achieve comparable reliability and validity metrics.

Summary of reliability and validity findings

Reliability: How consistent are the results of the IAR-DST across different settings and different users?

Reliability helps improve the quality of decisions made using a decision-support tool, and builds confidence in the tool's utility, making users more willing to rely on its recommendations. Reliable measures provide consistent data, which allows decision-makers to have faith in the information they're using. Inter-rater reliability refers to the degree of agreement or concordance between two independent raters; it tells you if the instrument can be reliably administered by different people to reach the same result, which is important for a tool which is used by many different raters across different settings.

We tested reliability of the IAR-DST by: reviewing all IAR-DST ratings available in the PMHC-MDS, and examining patterns of usage across Primary Health Networks (PHNs), conducting an observational study of IAR-DST usage at a Medicare Mental Health Centre, and by conducting an independent study of 100 people completing the IAR-DST and having different raters rate their level of care.

Key Findings:

Concordance rates between the IAR-DST recommended Level of Care and the practitioner-assigned Level of Care nationally (N=44658) are very high (89%).

- Observation data collected during standard process at an adult mental health clinic also demonstrated there was a **very high level of concordance (91%) between the level assigned by intake officers and the level assigned by clinical staff at review.**
- **Tool usage and reporting practices appear to vary significantly between PHNs**. For example, the rates of upgrades and downgrades in levels of care vary greatly across PHNs, suggesting some PHNs are using the tool to access a particular care pathway, and completeness of the data reported up to the PMHC-MDS also varies across PHNs.
- **Practitioner-assigned Level of Care changes contribute to the high proportion of Level 3 ratings.** Almost half of all IAR-DST Level 1 recommendations are upgraded by the practitioner to a higher Level of Care, and over a quarter of Level 2 recommendations are upgraded to a Level 3 in a general 'movement to the middle'. This arises when the practitioner overrides the tool, suggesting that there is something about the tool in practice that is not serving the setting; it could be that levels are not precisely enough aligned with clinical need, or it may be a result of a limited number of local referral options motivating raters to up or down grade to access a service.
- **Inter-rater agreement varies according to the complexity of the domain.** There is higher inter-rater agreement for Domain 2, Harm–a relatively straightforward and objective domain–than for Domains 6, 7, and 8, which are more complex and subjective. Overall, inter-rater agreement was moderate for the primary domains.

But a test or instrument can be reliable without being valid. Consistent results do not always indicate that a test is measuring what it is intended to measure.

Validity: Do the IAR-DST Domain scores consistently relate to mental health support needs?

A test is valid if it measures what it claims to measure. What it claims to measure is the construct of interest – in the case of the IAR-DST, the construct is mental health support needs. Construct validity is the degree to which the tool accurately measures the construct of interest. One test of construct validity is convergent validity – an assessment of the degree to which two measures that theoretically should be related (i.e. because they claim to measure the same construct), are related. If the scores show a positive correlation, this indicates they are measuring the same construct.

We assessed convergent validity by comparing IAR-DST Domain scores to a battery of standardised, well-validated, self-reported mental health resources.

Key Findings:

- **Convergent validity between IAR-DST Domains and self-report measures appears to vary according to domain complexity.** Convergent validity was highest for Domain 2: Harm and lowest for Domain 6: Social and environmental stressors.
- Use of a total score from a well-validated self-report measure (averaged across a larger number of survey questions) generally showed higher convergent validity with Domain scores than matching fewer items to the Domain content.
- **Use of standardised self-report measures to co-assess IAR-DST Domains could improve the construct validity of the tool.** This could be implemented with a self-report component to the IAR-DST, which could have the added benefit of involving the client more in the provision of details about their concerns and symptoms, and as a result enhancing the experience of a shared decision-making encounter.
- **Use of self-report data to increase the construct validity of the tool could be targeted to the more complex Domains that measure multiple constructs** (Domains 1, 3, 4 and 6). Further testing would be needed to determine the best self-report instruments (with the least burden) to enhance coverage of the broad spectrum of symptoms and features covered in the more complex domains.
- **Careful consideration to balance client burden and tool accuracy will be needed.** The addition of any self-report instruments to the process should not compromise the positive aspects of the IAR-DST interview experience as reported in these studies – participants reported a positive experience of the interview which they perceived as respectful, attuned to their needs and holistic.

Validity: Does the IAR-DST relate to an existing 'gold standard' measure?

Criterion validity is the degree to which the tool is an adequate reflection of a 'gold standard'. One method to test criterion validity is concurrent validity, this is when a new measure is compared to a well-validated existing measure that is administered at the same time. Another is predictive validity, which is typically a regression analysis to predict the criterion variable based on the test scores.

Using the PMHC-MDS we tested the IAR-DST Levels of Care against concurrently collected K-10+ levels of distress scores.

The Kessler Psychological Distress Scale Plus (K10+) is a 14-item screening tool designed to identify significant psychological distress and its functional impact over the past 4-weeks.

Key Findings:

- **There is a significant, but weak, positive relationship between the IAR-DST Level of Care and concurrent K10+ level of distress scores (PMHC-MDS; N= 27416).**
- **Linear regression results show there is a reliable predictive relationship; as IAR-DST Level of Care increases, K10+ level of distress scores increase, but the effect size is small.**
- **There is a similarly reliable, but weak, positive relationship between IAR-DST Domain 1 Symptom Severity and Distress scores and K10+ level of distress scores; as Domain 1 scores increase, K10+ scores increase.**
- **In summary, the symptoms measured by the K10+ and those assessed by the IAR-DST do vary in the same direction as each other, but they do not closely align with each other. This may be a reflection of the tool, but equally it could indicate that the K10+ is limited as a measure to discriminate between levels of severity in mental illness. It is beyond the scope of the data in this study to draw stronger conclusions.**

Key Recommendation:

- **Evaluating the IAR against broader and more sensitive measures of mental ill health is advisable, such as those included in the battery in Rapid Validity Study 2.**

Validity: Taken at face value, does the IAR-DST appear to test what it aims to test?

Face validity is whether the test appears (at face value) to measure what it claims to. Although face validity is a rather subjective assessment, it is nonetheless an important part of content validity because the initial impressions and intuitive understandings of the tool by the target population can significantly influence participation rates and the quality of responses.

For the IAR-DST, the target populations who are well-placed to judge its face validity are people who work with the tool, and individuals who are interviewed and classified according to the tool.

Key Findings:

- **Face validity appears to be high among clients**; interview participants in Rapid Study 2 considered the IAR-DST to cover all areas of mental health support needs.
- **Face validity is mixed among health professionals who use the tool.** Face validity appears higher for clinicians who receive referrals with an IAR-DST level of care than those who use the tool to assign a level of care, and lowest amongst GPs.
- However, **face validity is mixed within GPs and appears to vary according to level of experience.** The majority of GPs who provided feedback on the tool using the direct feedback option online (N=143) reported they found the tool to be generally helpful–specifically that the IAR-DST was helpful in understanding the initial assessment and making an appropriate decision.

Key Recommendation:

More work needs to be done to make the IAR-DST acceptable to GPs.

- **Offering GPs a facility to complete the IAR-DST which acknowledges their clinical expertise** and recognises that it will be more useful for some than others, could improve tool acceptability and uptake.
- **Integration into GP software and treatment and support databases is needed**, this would streamline the process of using the tool for the busy GP and provide them more readily with the outcome from the tool that they are looking for in their brief encounter with a client, which is to offer them access to a service.
- **Updates or improvements to the tool should take into account the current high face validity for clients.** People like the holistic approach, it makes them feel heard, and this is an important success factor at the front door of our mental health system.

Validity: Does the content of the IAR-DST adequately reflect the construct being tested?

Content validity is the degree to which the content is an adequate reflection of the construct measured. Generally, experts on the subject matter determine whether or not a test has acceptable content validity.

Our rapid desktop review concluded that there was adequate evidence of content validity from subject-matter experts, however, this could be strengthened by further feedback on the IAR-DST in the contexts where it is currently used.

Key Findings:

Content validity was strengthened by the feedback from people with mental health concerns interviewed in Rapid Study 2. Feedback was broadly positive, but specifically:

- Interviewees reported that **the questions asked during the interview covered all areas of their support needs.**
- Interviewees **appreciated having clear questions and examples provided**, which further supports the content validity of the Domains, and more specifically, the Domain guidance notes.

Content validity was challenged by the feedback from health professionals who used the IAR-DST in Rapid Study 3. Feedback was mixed, but specifically:

- Positively, some health professionals stated that the IAR-DST does provide clear common language to speak with other clinicians.
- Negatively, some health professionals reported there are areas that are not addressed or included in the Domain guidance notes (eg. disorders that do not present as acute/complex, however, require specialist services), indicating more work needs to be done to improve (at least) the face validity of Domain coverage, but also to strengthen content validity.

IAR Validity project activities:
Results

Activity 1:

Rapid desktop review of existing IAR-DST evaluations and reviews

Overall, to date, there has been no rigorous, planned or consistent evaluation of validity and reliability of the IAR-DST as a standalone tool, or within the multiple settings where it is now used.

However, there have been a number of activities, where data has been collected that speaks to aspects of validity. For the most part these activities have related to two aspects of validity –face validity (at a glance, does the question seem to measure what we want it to measure?) and content validity (examination of theoretically whether the question could measure the construct of interest). The activities have employed different approaches and methodologies, which make comparison and overall conclusions more difficult.

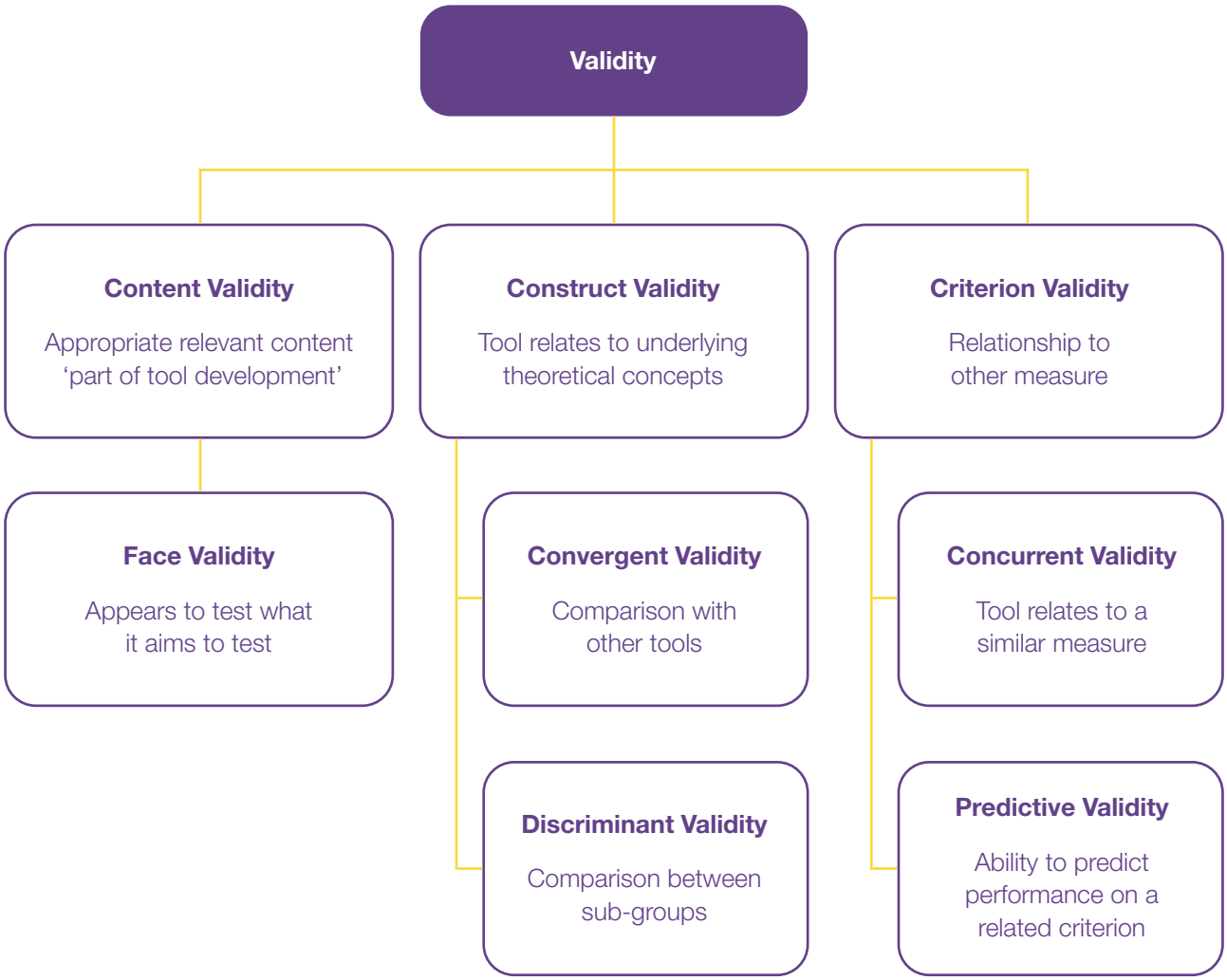
Importantly, there is very little data to allow evaluation of the most vital aspects of validity and reliability for a decision support tool – including the tool's alignment with other accepted ways of measuring mental health clinical need such as validated self-report measures, the tool's ability to discriminate between groups and presentations, and the tool's ability to predict key outcomes (e.g. Level of Care needed, or outcome that can be expected or achieved).

Evidence of past validation activities has been summarised in the IAR Review Project Interim Insights Report (NOUS; 2023) with a traffic-light coding system. Face and content validity are assessed as tested (green) through expert opinion, The University of Melbourne (UoM) Implementation Review, training surveys and public consultation. Construct validity is assessed as partially tested (orange) through the UoM Implementation review and research conducted by a South Australian Local Hospital Network. Criterion validity – which includes the tool's alignment with other standard measures of mental health clinical need and the tool's ability to predict outcomes – are assessed as not tested (red). To better understand what further validity evidence is required, we reviewed all available documents to further evaluate the quality of the suggested evidence of past validation.

For **content and face validity**, the current evidence comes from the Expert Advisory Group (EAG) (APS 2021; DoHAC, March 2022; DoHAC, May 2024), Literature review (APS, 2018), and Implementation review (UoM, 2021). For **construct validity**, the main evidence comes from the UoM implementation review. Information about the evaluation of the evidence for content, face and construct validity is presented in Table 1.

Our evaluation of the available evidence of validity of the IAR-DST was guided by the COSMIN rating tool (Consensus-based Standards for the Selection of health Measurement Instruments; Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018; see Figure 1) which is a systematic tool for appraising and comparing the quality and validity components of a measure. COSMIN defines **validity** as the degree to which an instrument measures the construct (psychological or behavioural phenomenon) it purports to measure.

Figure 1. Types of validity as defined by COSMIN



Activity 1: **Rapid desktop review continued.**

Table 1. Evaluation of IAR-DST review activities to date and any evidence of validity they provide

	COSMIN definition	Quality checklist	Expert Advisory Group	APS Literature Review	UoM Implement. Review	OVERALL RATING
Content validity	The degree to which the content is an adequate reflection of the construct measured	Patients asked?	✓	✗	✓	**PARTIAL EVIDENCE** **WORK TO DO: TESTING NEEDED**
		Professionals asked?	✓	✗	✓	
		Relevance?	✓	✓ (some evidence)	✓ (some evidence)	
		Comprehensiveness	✗	✓ (some evidence)	✓ (some evidence)	
		Comprehensibility?	✗	✗	✗	
		Including items?	✓	✗	✓	
		Including responses? (rating point descriptors)	✓	✗	✗	
		Including instructions? (glossary)	✓	✗	✗	
Face validity (part of content validity)	The degree to which a measure appears to be an adequate reflection of the construct	8 domains of IAR-DST *appear* relevant	✓	✓	✓	**GOOD ENOUGH DATA**
		Appears representative of mental health treatment need	✓	✓	✓	
Construct validity	The degree to which the tool consistently measures the construct of interest	Convergent validity: comparison with other measures	✗	✗	✗	**NO DATA**
		Discriminative: comparison between sub-groups	✗	✗	✗	
Criterion validity	The degree to which the tool is an adequate reflection of a 'gold standard'	Concurrent: relates to a similar measure	✗	✗	✗	**NO DATA**
		Predictive: predicts performance on a related criterion	✗	✗	✗	

Activity 1: **Rapid desktop review continued.**

The Expert Advisory Group (EAG) has involved over 40 members with expertise in mental health, including people with lived experience and carers, since 2017. We reviewed the available evidence of EAG involvement in tasks that appear to demonstrate contributions to content and face validity (APS 2021; DoHAC, March 2022; DoHAC, May 2024).

COSMIN requires that to assess content validity, an adequate sample of patients and professionals need to be asked about the relevance, comprehensiveness (completeness including everything that is necessary) and comprehensibility (ease of understanding) of the items, response options and instructions.

Our review finds that there is evidence of a contribution to content and face validity in the following tasks undertaken by the EAG:

1. The EAG updated descriptions of domains, rating points and services to include content relevant to each cohort (adults, children, adolescents, older adults)
2. The EAG suggested updates to terminology in descriptions to be more relevant and accurate (e.g. to differentiate children's development from mental health concerns)
3. The EAG reviewed the IAR-DST guidance and provided comprehensive feedback on content to ensure it is clinically useful
4. The EAG reviewed the IAR-DST to ensure it is person-centred, respectful of, and responsive to the preferences and needs of people with lived experience of mental health concerns

The **systematic literature review**, commissioned by the Department of Health, Disability and Ageing and **produced by the Australian Psychological Society in 2017-2018 and updated in 2021** (APS, 2018), included evidence from 13 international stepped care approaches in primary mental health which suggested important points for further consideration. From a clinical utility perspective, this review highlighted the importance of matching people's needs with the most appropriate step for intervention as well as establishing solid processes between step-ups and step-downs between levels.

We reviewed the available evidence of content and face validity provided in the review. Specifically, we highlight that content and face validity is evidenced through the following suggestions emerging from the review:

- Assess multifaceted nature of risks appropriately (e.g. family violence, suicide, self-harm, harm to others) and timely (monitor at every visit)
- Qualified professional to assess risks, comorbidity, age, gender, culture, and person's current circumstances and context

Activity 1: **Rapid desktop review continued.**

The implementation review, facilitated by the Department, and completed by **The University of Melbourne** (UoM, 2021), included a review of 9 Primary Health Networks (PHNs) to better understand barriers and enablers to implementation of the IAR-DST by conducting a stakeholder survey and workshops, where 8 key recommendations were produced. The review highlighted the importance of clearly distinguishing between levels of care and providing clear guidance on step-up and step-down strategies and considering the reframing of 'initial' assessment as patients often have multiple visits and need to retell their health history. The authors found that the prioritisation of primary domains is consistent with the decision logic and that the IAR-DST supports linking assessment results to a Level of Care.

The current review identified recommendations related to content and face validity; there was no evidence found for construct validity. Specifically, there is evidence of recommendations that could contribute to content and face validity in the following tasks undertaken by the implementation review:

- The review recommended the IAR-DST content needs to apply to wider populations (e.g. Aboriginal and Torres Strait Islander peoples, Culturally and Linguistically Diverse people, people with complex physical, social and mental health multimorbidity and people in acute care settings)
- The review recommended it is important to tailor the IAR-DST questions to different populations
- The review found the domains appear to be valid, acceptable and important to assess a person's mental health needs.

Conclusion

Based on this rapid review we proposed the following set of studies to provide further evidence of reliability and validity (Table 2, below):

Table 2. Timeline of Activities and Evidence of Validation

Timeline	Activity	Tasks that constitute evidence of validation			
		Reliability (inter-rater*)	**Content & face validity**	**Construct validity**	**Criterion validity**
2024-2025	Observation study at Medicare Mental Health Centre	✓	✓ (face validity)	✓	✓
2024-2025	Directly recruited representative sample who will complete the IAR-DST with InsideOut Institute staff	✓	✓	✓	✓
2024-2025	Survey for clinicians using IAR-DST	N/A	✓ (face validity)	N/A	N/A
2024-2025	Analysis of PMHC-MDS data	✓	N/A	✓	✓

* ***Inter-rater reliability*** is the extent to which scores for patients who have not changed are the same for repeated measurement when assessed by different persons (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). Other measures of reliability (intra-rater, test-retest, internal consistency) are less applicable.

Activity 2:

Validity and reliability of the IAR-DST from existing data sources

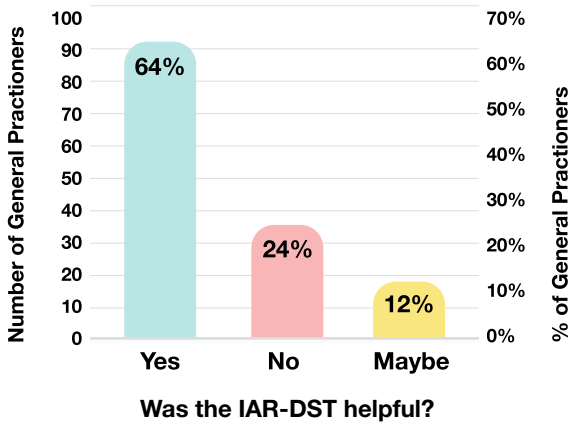
1. The IAR-DST Direct feedback form

Finding: Majority of GPs found the IAR-DST to be helpful

We reviewed the results of the **online feedback form from 2022 to end of May 2024**. Of 228 valid respondents, 63% (n=143) were general practitioners (GPs)(Fig. 2). For these 143 respondents, we examined responses to the question about whether they found the IAR-DST helpful.

Q. ***Was the decision support tool helpful in undertaking the initial assessment and making a decision about an appropriate Level of Care?***

Figure 2. Online feedback responses from GPs (n=143)



Two-thirds of GPs said that the IAR-DST was helpful in understanding the initial assessment and making an appropriate decision about the Level of Care.

Less than a quarter of GPs said the tool was not helpful. The majority of those who were undecided stated that they would need more experience before they could comment.

We then examined the feedback regarding recommended changes only from the respondents who endorsed the tool or were as-yet undecided.

Q. ***What changes do you recommend to the decision support tool and/or the Initial Assessment and Referral Guidance?***

The recommendations from this cohort were mostly technical and referred to increasing utility through integration of the tool and streamlining the user experience. Only a few respondents had more substantive recommendations, such as reducing the number of domains, or changing details in the guidance notes.

Key recommendations from feedback:

Integration with practice software, referrals and link to services (17 recommendations)

Improve user experience: add pop-ups or hovers for guidance notes to reduce screen switching, shorten the guidance notes, improve saving or printing (13 recommendations)

Activity 2: **Validity and reliability of the IAR-DST from existing data sources continued.**

2. Results from the PMHC-MDS aggregated reports

Analyses were performed on aggregate data extracted October 2024 for the period 2020-2024.

Analysis of the PMHC-MDC data provides results on concordance between the IAR-DST recommended and practitioner-assigned Level of Care across and between PHNs.

Overall, the review of the data was positive. Concordance between the IAR-DST recommended Level of Care and practitioner-assigned Level of Care was very high (89%).

However, by further aggregating the incidence of level changes (upgrades and downgrades to the IAR-DST recommended Level of Care by the practitioner), it is apparent that a significant number of Level 1 and 2 recommendations are being upgraded. This impedes the utility of the tool to contribute to the aims of stepped care and referral to low-intensity interventions. Comparison between PHNs shows significant differences in reporting and usage of the tool which warrant further investigation.

Analysis across PHNs

Across all PHNs which reported both the IAR-DST recommended level and the practitioner Level of Care recommended, we examined concordance, and the incidence of level changes.

Overall, the concordance rates between the IAR-DST recommended Level of Care and the practitioner-recommended Level of Care are very high (89%). To further investigate the incidence of level changes, we aggregated the level changes across all PHNs to view rate-change trends.

Table 3. IAR-DST-recommended level of care vs. practitioner-assigned level across PHNs

IAR-DST	Practitioner-assigned Level of Care					TOTAL
	Level 1	Level 2	Level 3	Level 4	Level 5	
1	587	478 ↑	53 ↑↑			1118
2	249 ↓	2513	1037 ↑			3799
3	54 ↓↓	1150 ↓	30834	606 ↑		32644
4			1278 ↓	5532		6810
5			53 ↓↓	38 ↓	196	287
TOTAL	**890**	**4141**	**33255**	**6176**	**196**	**44658***

* ***Total practitioner-assigned level of care*** includes all primary mental health care program types. Completion of the IAR-DST is only compulsory for Medicare Mental Health Centres/ Head to Health Centres. Rates of IAR-DST completion vary across programs and therefore distributions may not be representative for all programs. 5% of overall reported usage does not state recommended Level of Care so not included in total.

Finding: By aggregating data across all PHNs, we can see that almost half of all IAR-DST Level 1 recommendations are upgraded by the practitioner to a higher Level of Care, and over a quarter of Level 2 recommendations are upgraded to a level 3.

Activity 2: **Validity and reliability of the IAR-DST from existing data sources continued.**

2. Results from the PMHC-MDS aggregated reports continued

Specifically:

- **only 53% of IAR-DST Level 1 recommendations are also assigned a Level 1 by the practitioner**, 43% are upgraded to a Level 2, and 4% are upgraded to a Level 3.
- 66% of IAR-DST Level 2 recommendations are also assigned a Level 2 by the practitioner, 7% are downgraded to a Level 1, and 27% are upgraded to a Level 3.
- **94% of IAR-DST Level 3 recommendations are also assigned a Level 3 by the practitioner**, 4% are downgraded to a Level 2, 2% are upgraded to a Level 4.
- 81% of IAR-DST Level 4 recommendations are also assigned a Level 4 by the practitioner, 19% are downgraded to a Level 3.
- 68% of IAR-DST Level 5 recommendations are also assigned a Level 5 by the practitioner, 13% are downgraded to a Level 4, and 18% are downgraded to a Level 3.

To further examine the drivers of these trends, we examine differences between PHNs.

Analysis between PHNs



Finding: Analysis suggests significant differences in usage between PHNs.

IAR-DST level versus practitioner-recommended level: Differences between PHNs

The incidence of upgrades and downgrades varies by PHN, strongly indicating differential use of the IAR-DST by location. For the majority of PHNs, there was a relatively normal rate of upgrades and downgrades as would be expected from the overall rates illustrated in the preceding data. However, there were outliers, which are important to consider.

In one PHN 57% of IAR-DST Level 1 recommendations were upgraded to a level 2 by the practitioner, and in another, 66% of Level 5 recommendations were downgraded to a Level 4 or a Level 3.

At the very ends of the outlier spectrum, one PHN had no incidence of upgrades or downgrades at all, while others displayed very high rates of change. In one PHN, 77% of IAR-DST Level 2 recommendations were downgraded to a Level 1 by the practitioner, and 74% of Level 3 recommendations were downgraded to a Level 2 or Level 1. In another PHN, 91% of Level 3 recommendations were downgraded to a Level 2 or Level 1 by the practitioner, and all (100%) Level 2 recommendations were downgraded to a Level 1.

Although the two PHNs have relatively lower usage of the tool than other PHNs, these very high rates of change are a driver of some of the overall rates of level change noted in the data aggregated across all PHNs. These very low rates of concordance between the IAR-DST recommended level and the practitioner-assigned Level of Care at these PHNs warrant further investigation, but perhaps reflect users working in a setting where care options are limited, modifying the assigned level of care to gain access to the available care options for consumers.

Activity 2: **Validity and reliability of the IAR-DST from existing data sources continued.**

3. Further results from the PMHC-MDS

Analyses were performed on unit-level data extracted November 2024 for the period 2020-2024.

Analysis of the PMHC-MDS unit-level data provides results on concordance between the IAR-DST recommended Level of Care and K-10 level of distress.

For a subset of service contacts (N=27,416) there was an IAR-DST at intake linked to an episode where the K10+ was collected. For a subset of these, full demographic information was also able to be linked (N=22,839, Table 4). Demographic information below provides a representative sample for the following analyses. The majority of the sample was born in Australia or New Zealand, around 60% female, more than half 36 years or older, the majority were assigned an IAR-DST Level 3 at intake, and almost half had a K10+ score indicating very high levels of distress.

Table 4. Subset of service contacts with linked IAR-DST, K10+ and demographic information (N=22,839)*

Gender	**N**	**%**
Male	8314	36.4
Female	14062	61.6
Other	342	1.5
Not stated	121	0.5
Age		
<12 years	110	0.5
12-15 years	739	3.2
16-17 years	745	3.3
18-25 years	4056	17.8
26-35 years	4936	21.6
36-49 years	5421	23.7
50-64 years	4836	21.2
65+ years	1992	8.7
Country of Birth		
Australia & NZ	19855	86.9
Other	2984	13.1

Remoteness of residence**		
Major Cities	10765	47.1
Inner Regional	9913	43.4
Outer regional-remote	2161	9.5
SEIFA disadvantage***		
1 (most disadvantage)	3177	13.9
2	6118	26.8
3	6387	28
4	4437	19.4
5 (least disadvantage)	2720	11.9
K10+ score level		
Low	2565	11.2
Moderate	3261	14.3
High	5900	25.8
Very High	1113	48.7
IAR-DST Level of Care		
Level 1	839	3.7
Level 2	2222	9.7
Level 3	16425	71.9
Level 4	3223	14.1
Level 5	130	0.6

* ***The PMHC-MDS*** identifies service contacts not unique clients. Clients may be counted more than once if they have service contacts with multiple providers.

** ***Remoteness of residence*** is derived using the Australian Statistical Geography Standard (ASGS) Remoteness Structure. Remoteness Areas divide Australia into five classes based on relative access to services: Major Cities, Inner Regional, Outer Regional, Remote and Very Remote. Client postcodes are used to determine Remoteness of Residence.

*** ***Socio-economic Indexes for Areas (SEIFA)*** Disadvantage ranks geographical areas according to their relative socio-economic advantage and disadvantage using Census data. Lower scores indicate higher relative disadvantage.

Mapping IAR-DST Level of Care to K10+ level of distress

Our prediction was that as IAR-DST Level of Care increased, K10+ level of distress would increase. Using K10+ total score categories set by the Australian Bureau of Statistics (ABS), we proposed the following direct mapping between IAR-DST Level of Care to K10+ level of distress (Table 5) and tested whether there was a significant relationship on the subset with a linked IAR-DST, K10+ (N=27,416).

Table 5. Direct mapping between IAR-DST Level of Care to K10+ level of distress tested

IAR Level of Care	K10+ total score
Level 1	K10+ Low (10-15)
Level 2	K10+ Moderate (16-21)
Level 3	K10+ High (22-29)
Level 4 or 5	K10+ Very High (30-50)

Relationship between IAR-DST Level of Care to K10+ level of distress

To explore whether there was a relationship between IAR-DST levels of care and K10+ level of distress, we cross-tabulated K10+ total score categories by IAR-DST Level of Care recommended (Table 6).

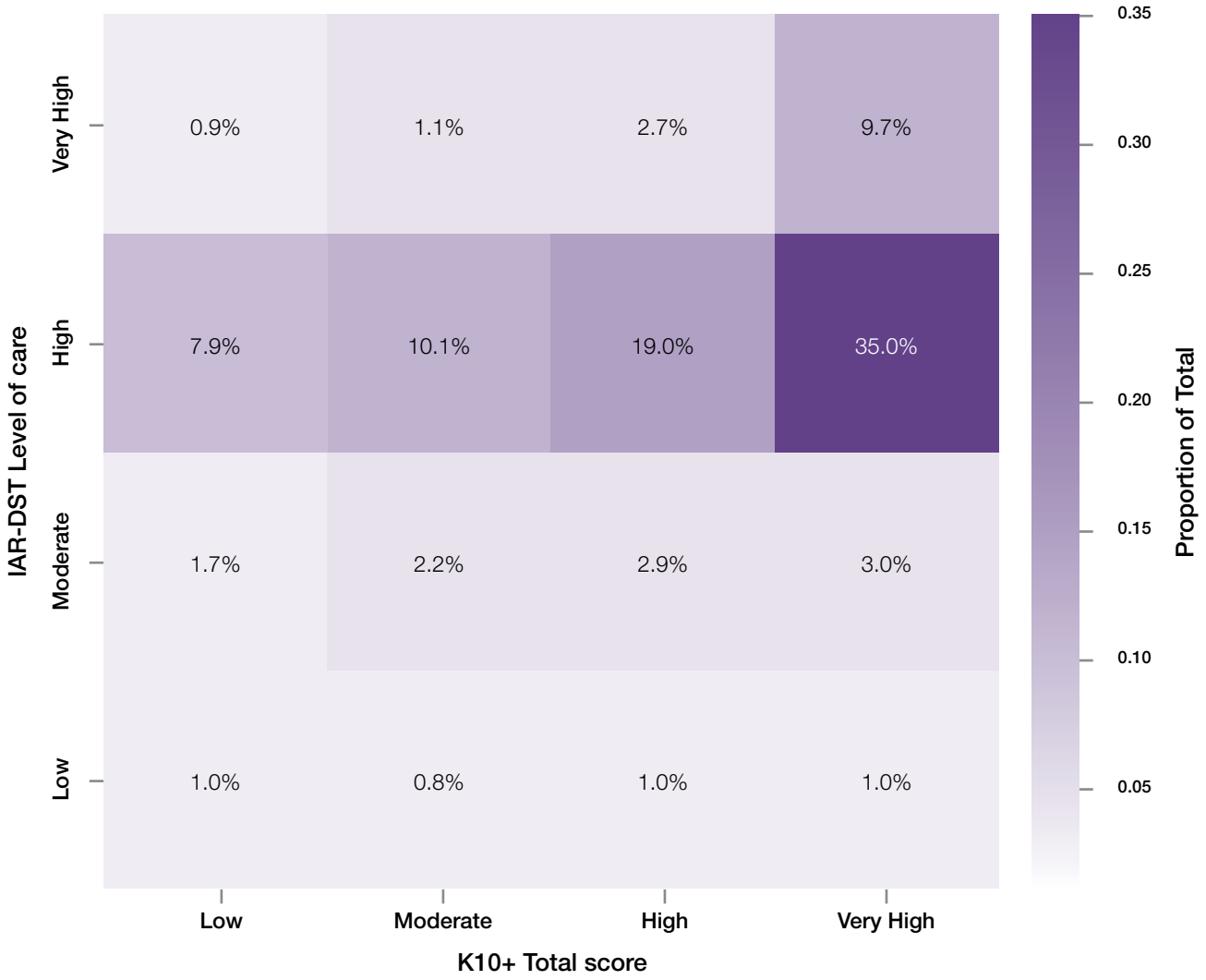
The majority of the sample were assigned a Level 3 or above and their K10+ score indicated a very high level of distress, as visually depicted as cross-tabulated proportions in Figure 3.

Table 6. Cross tabulated frequencies mapping IAR-DST Level of Care to K10+ level of distress

IAR-DST	K10+ Level of distress				
	Low	Moderate	High	Very High	Total
Level 1	272	219	277	262	1030
Level 2	459	590	800	824	2673
Level 3	2171	2773	5200	9605	19749
Level 4+5	244	313	749	2658	3964
Total	**3146**	**3895**	**7026**	**13349**	**27416**

Mapping IAR-DST Level of Care to K10+ level of distress continued.

Figure 3. Heatmap: Cross-tabulated IAR-DST Level of Care to K10+ level of distress proportions (N= 27416)



Relationship between IAR-DST Level of Care to K10+ level of distress: Statistical tests

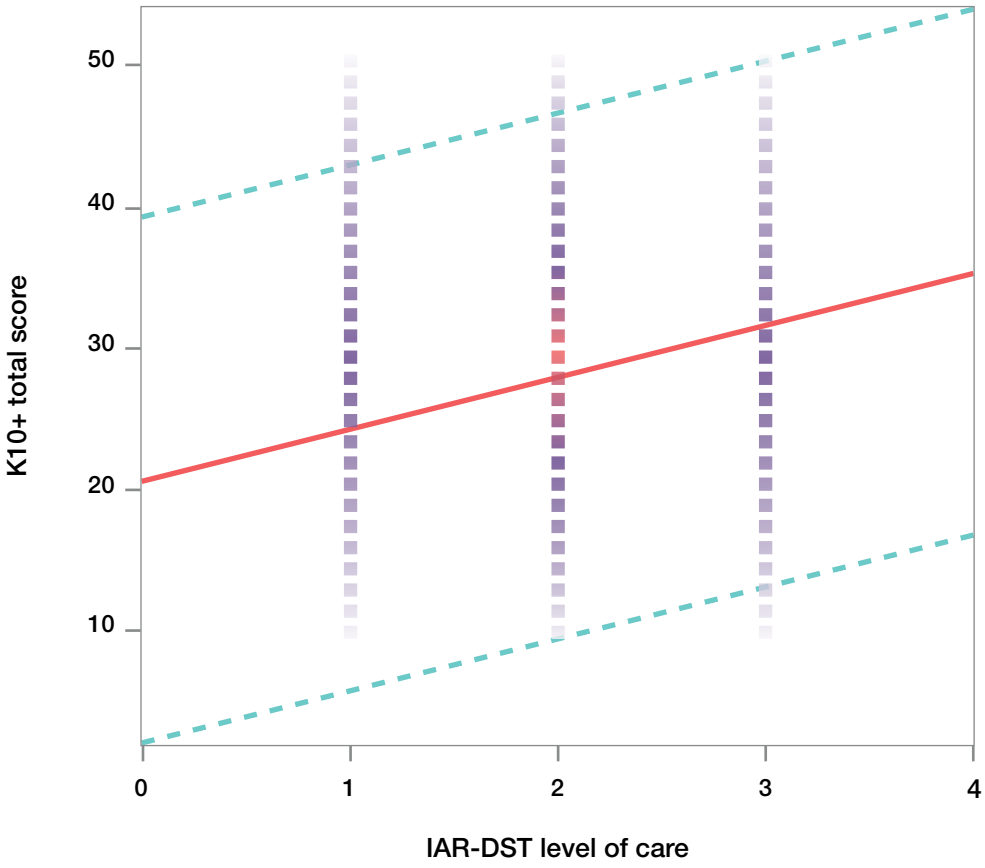
An initial chi-squared test demonstrated there was a significant relationship between IAR-DST Level of Care and K10+ level of distress X^2 (6, N=27,416) = 10444.84, p= <0.0001.

To explore the relationship, we tested the exact agreement between IAR-DST Level of Care and K10+ level of distress based on the proposed direct mapping using linear weighted kappa. The kappa value was K = 0.11, SE=0.0036, CI [0.1051, 0.1191] indicating a significant but low agreement (K < 0.20 = slight agreement). Spearman correlations also indicated a significant, but weak, positive relationship r = 0.21, p <0.0001.

Mapping IAR-DST Level of Care to K10+ level of distress continued.

To further assess the relationship, a linear regression was conducted to evaluate the extent to which Level of Care predicts K10+ level of distress. The model was significant, but Level of Care explained only a small portion of the variance in K10+ scores ($F(1, 27414) = 1513.14$ $p < 0.0001$, $R^2 = 0.05$). The linear trend is displayed in Figure 4. The statistically significant but low regression slope indicates there is a reliable positive relationship as predicted –as IAR-DST Level of Care increases, K10+ level of distress scores increase–but the effect size is small.

Figure 4. Linear regression slope predicting K10+ score from IAR-DST Level of Care (N= 27416)



Key Finding: There is a reliable, but weak, positive relationship between IAR-DST Level of Care and K10+ level of distress scores–as Levels of Care increase, K10+ scores increase.

Mapping K10+ scores to IAR-DST Domain 1 symptom severity and distress scores

IAR-DST Domain 1 symptom severity and distress scores are more closely related to K10+ level of distress scores; therefore, we also assessed the relationship between these variables. We proposed the following direct mapping between IAR-DST Domain 1 symptom severity and distress scores to K10+ level of distress and tested whether there was a significant relationship between the variables (Table 7).

Table 7. Direct mapping between IAR-DST Domain 1 to K10+ level of distress tested

IAR Domain 1 score interpretations	K10+ total score
0= No problem	K10+ Low (10-15)
1= Mild distress 2= Mild to moderate distress	K10+ Moderate (16-21)
3= Severe distress most of the time	K10+ High (22-29)
4= Severe distress all the time	K10+ Very High (30-50)

Relationship between IAR-DST Domain 1 scores and K10+ level of distress

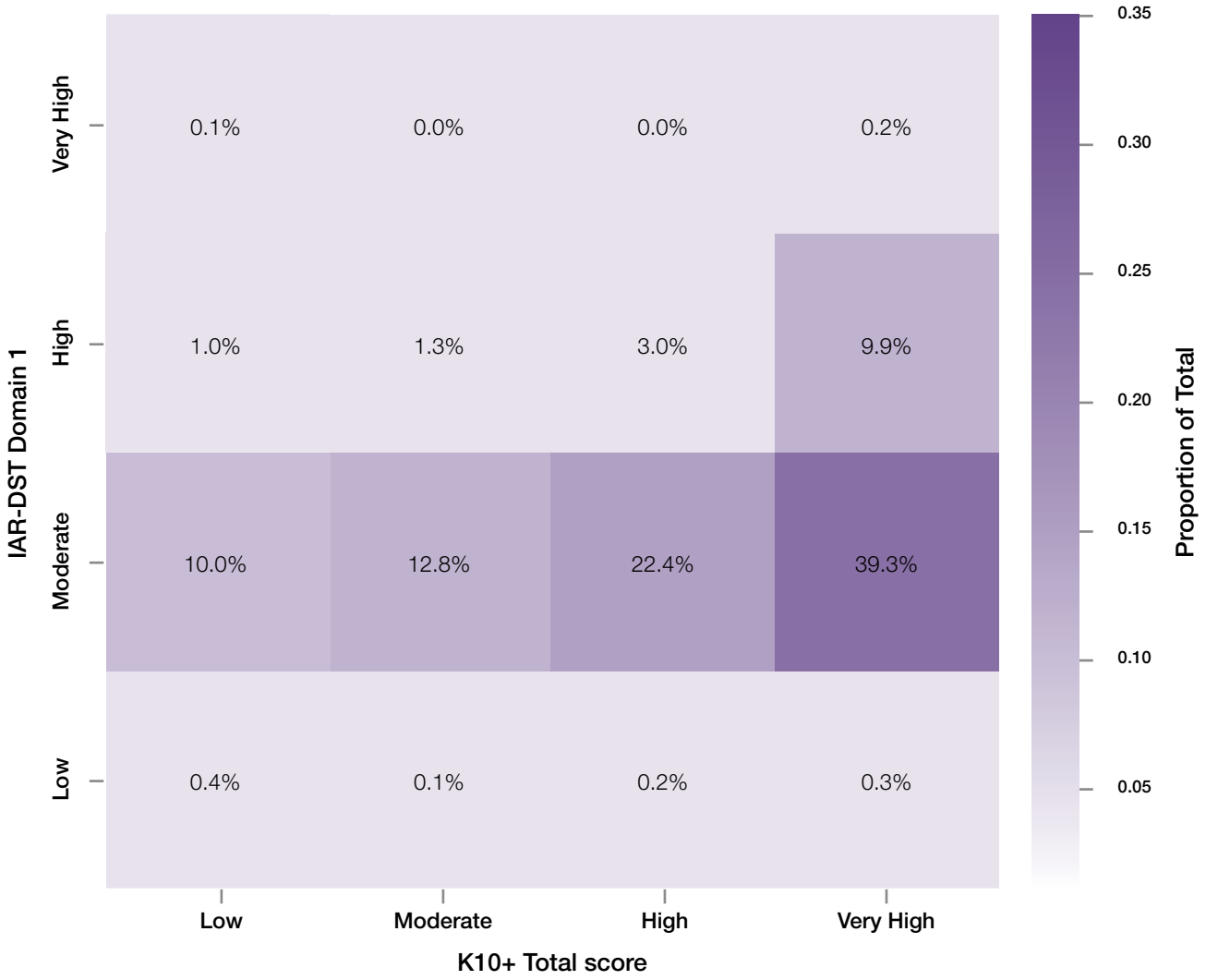
To explore whether there was a relationship between IAR-DST Domain 1 scores and K10+ level of distress, we cross-tabulated K10+ total score categories by IAR-DST symptom severity and distress scores (Table 8). The majority of the sample were assigned Domain 1 score of 1 or 2, indicating mild or mild to moderate distress and a K10+ score indicating a very high level of distress, as visually depicted as cross-tabulated proportions in Figure 5.

Table 8. Cross tabulated frequencies mapping IAR-DST Domain 1 symptom severity and distress score to K10+ level of distress

IAR-DST Domain 1	K10+ Level of distress				
	Low	Moderate	High	Very High	Total
0 = No problem	112	35	47	73	267
1 = Mild distress 2 = Mild to moderate distress	2751	3498	6148	10510	22907
3 = Severe distress most of the time	267	355	823	2711	4156
4 = Severe distress all the time	16	7	8	55	86
Total	**3146**	**3895**	**7026**	**13349**	**27416**

Mapping K10+ scores to IAR-DST Domain 1 symptom severity and distress scores continued.

Figure 5. Heatmap: Cross-tabulated IAR-DST Domain 1 scores to K10+ level of distress proportions (N= 27416)



Relationship between Domain 1 scores and K10+ level of distress: Statistical tests

An initial chi-squared test demonstrated there was a significant relationship between IAR-DST Domain 1 symptom severity and distress score and K10+ level of distress X^2 (6, N=27,416)=21175.04, p=<0.0001.

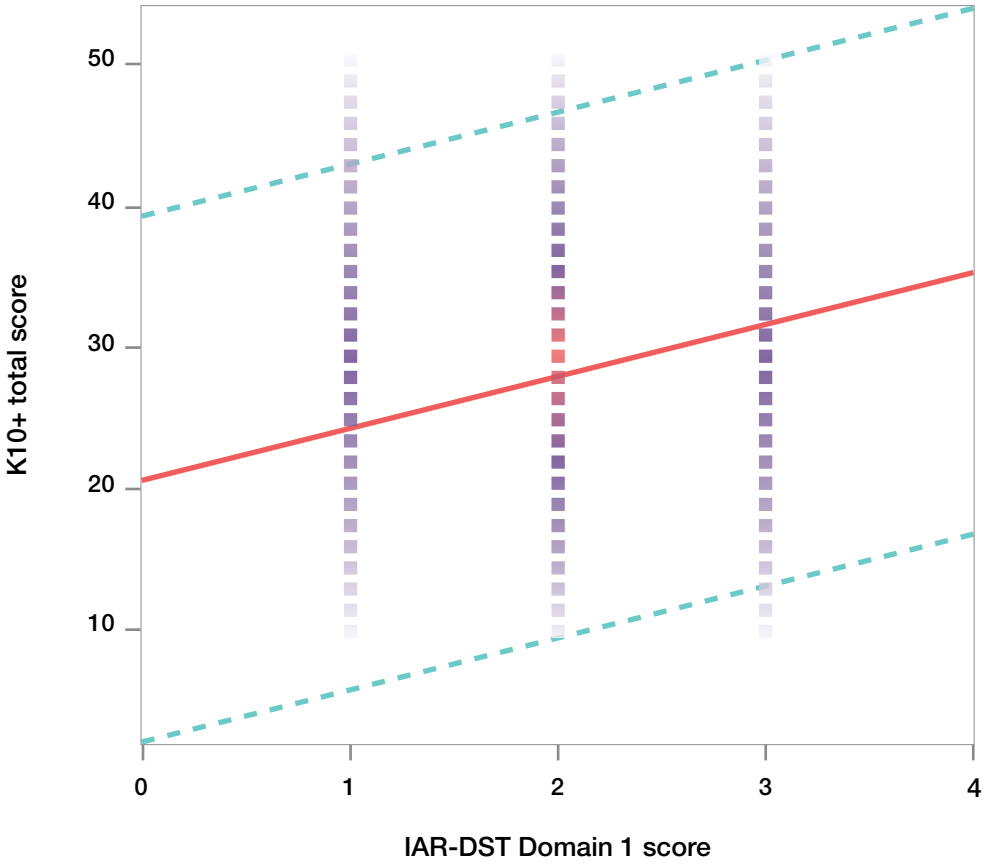
To explore the relationship, we tested the exact agreement between IAR-DST Domain 1 symptom severity and distress score and K10+ level of distress based on the proposed direct mapping using linear weighted kappa. The kappa value was K=0.03, SE=0.0015, CI [0.0278, 0.0337], indicating a significant but low agreement (K < 0.20 = slight agreement). Spearman correlations also indicated a significant, but weak, positive relationship (r = 0.18, p <0.0001).

Mapping K10+ scores to IAR-DST Domain 1 symptom severity and distress scores continued.

To further assess the relationship, a linear regression was conducted to evaluate the extent to which Domain 1 scores predict K10+ level of distress. The model was significant, but Domain 1 scores explained only a small portion of the variance in K10+ scores ($F(1, 27414) = 1161.63$, $p<0.0001$, $R^2 = 0.04$).

The linear trend is displayed in Figure 6. The statistically significant but low regression slope indicates there is a reliable positive relationship; as IAR-DST Domain 1 scores increase, K10+ level of distress scores increase, but the effect size is small.

Figure 6. Linear regression predicting K10+ score from IAR-DST Domain 1 score (N= 27416)



Key Finding: There is a similarly reliable, but weak, positive relationship between IAR-DST Domain 1 Symptom Severity and Distress scores and K10+ level of distress score; as Domain 1 scores increase, K10+ scores increase.

Rapid Validity Studies

Rapid Study 1

Rapid study 1 was a 3-month observation study at a Medicare Mental Health Centre between 1st November 2024 – 31st January 2025.

Sociodemographic description of the sample

During the observation period, 238 adult clients aged 18-76 years (M=36 years) completed intake at the service (Table 9). More than half the clients identified as female (60%) and over half were assigned an IAR-DST Level 3 at intake (58%). Two-thirds of clients received either one or two referrals, and the most frequent referrals were to an adult mental health service (this usually means referred into the same service where the intake assessment was conducted).

Table 9. Sample demographics and clinical characteristics (N=238)

Gender	**N**	**%**
Male	84	35
Female	142	60
Other	10	4
Not stated	2	1
Age		
18 - 24	57	24
25 - 29	40	17
30 - 39	49	20.5
40 - 49	42	17.5
50 - 65	38	16
65+	12	5
IAR-DST Level of Care		
Level 1	4	1.5
Level 2	56	24
Level 3	139	58
Level 4 & 5	39	16.5

Most Common Referrals	**N**	**%**
Adult mental health service	110	21
Private allied health service	98	19
Community support groups / agencies	65	13
Other community / health care service	42	8
Telephone & online services / referral agency	41	8
Psychiatric / mental health service or facility	26	5
Peer support / self-help group	18	4
Community Based Drug and Alcohol Service	15	3
Total referrals / persons		
Level 1	72	30
Level 2	85	36
Level 3	51	21
Level 4 & 5	30	13

Concordance between intake and clinical review

As part of standard clinical practice, the IAR-DST levels of care recommended at intake are reviewed by the clinical team daily. A second rating by the clinical team was available for 221 IAR-DST levels of care. Of these, minimal changes were made; there were only 3 instances of level downgrades and 2 instances of level upgrades. There was therefore a very high level of concordance (91%) between the level recommended by intake officers and the level recommended by clinical staff.

Number of referrals by IAR-DST Level of Care

To explore whether there was a relationship between levels of care and number of referrals, we cross-tabulated number of referrals per person by IAR-DST Level of Care recommended (Table 10). There was no significant relationship between Level of Care and number of referrals X^2 (20, N=238) =13.51, p= 0.855.

Table 10. Number of referrals per person by IAR-DST Level of Care

IAR-DST	Number of referrals					
	1	**2**	**3**	**4**	**5**	**Total**
Level 1	1	2	1	0	0	4
Level 2	17	21	12	6	0	56
Level 3	36	50	32	18	3	139
Level 4	15	11	6	2	1	35
Level 5	3	1	0	0	0	4
Total	**72**	**85**	**51**	**26**	**4**	**238**



Key Finding: There was a very high level of concordance (91%) between the IAR-DST Level of Care assigned by intake officers and the level confirmed upon review by clinical staff

Rapid Study 2

Rapid study 2 was conducted with a sample of people who represent primary care patients between November 2024 and January 2025. Participants (n=100) undertook a comprehensive battery of standardised mental health assessments online, completed the IAR-DST via Zoom with an InsideOut Institute staff member trained as an intake officer, and completed a post-interview follow-up survey online.

Sociodemographic description of the sample

100 participants took part in this study. Almost three quarters identified as female (72%), aged between 18 and 63 with a mean age of 33.7 (SD=10.9). Respondents from every state and territory participated, the majority living in NSW (n=34) and Victoria (n=30). The majority were of Oceanian (64%) background, and the majority spoke English at home (90%). Half reported their relationship status as single, and most were in stable housing (83%). The majority (78%) reported living with others (roommate/ family/ a partner), and 22% reported living alone. Two-thirds reported an average income over $45k per annum, two-thirds were in full or part-time work and two-thirds were receiving government benefits.

Nearly half the cohort reported a major medical problem (n=48), half reported having no disability, and 38% reported a psychiatric disability. One third reported they were unable to carry out their usual activities for 1-5 days in the past month, one third reported they were unable to carry out usual activities for 6-30 days, and one third reported no functional impacts.

Mental health

Mental health or behavioural problems were reported by 80% of participants. A range of mental health conditions were reported, with high rates of depression and anxiety (both ~60%). Most participants had previously sought help previously from a psychologist (71%) or general practitioner (67%), and almost half had seen a psychiatrist (n=48). Most had found previous support at least somewhat helpful. A third had been hospitalised for mental health difficulties previously, and half were currently under treatment for mental health difficulties. Over half the participants had an immediate family member with a history of mental health difficulties.

Making changes to mental health

Participants rated four questions on a 5-point scale ranging from 1 (not at all) to 5 (extremely):

Importance to make changes to mental health, M=4.25 (SD=0.73)
Readiness to make changes to mental health, M=4.12 (SD=0.81)
Confidence to make changes to mental health M=3.37 (SD=1.07)
Motivation to work on current symptoms, M=3.89 (SD=0.90).

Key findings: A clinician rater consistently rated Levels of Care higher than a non-clinician intake officer, suggesting a rater's clinical experience and expertise is a significant factor affecting the reliability of the tool in real life settings.

All 100 participants were interviewed by InsideOut Institute researchers trained as intake officers, and reviewed by a clinician (clinical psychologist) who attended standard Primary Health Network training in the administration of the IAR-DST. Each participant completed an audio-recorded Zoom semi-structured interview lasting about 15-20 minutes asking questions about each of the domains that were composed by the research team based on the IAR-DST training and materials provided. The intake officer used the online IAR-DST to determine a recommended Level of Care and assign a Practitioner-assigned Level of Care. Interviews were rated by a second rater, a clinical psychologist, who listened to the recordings and completed the IAR-DST-recommended Level of Care and the Practitioner-assigned Level of Care.

Intake officer: Recommended vs. Assigned concordance

For the interviewer, the concordance between the IAR-DST recommended Level of Care and Practitioner-assigned Level of Care shows 85% agreement, with the interviewer downgrading 4 participants and upgrading 11 participants in terms of the assigned Levels of Care.

Clinician: Recommended vs. Assigned concordance

In the case of the clinician rater, the concordance between the IAR-DST recommended Level of Care and Practitioner-assigned Level of Care shows 83% agreement, with the clinician rater downgrading 6 participants and upgrading 10 participants in terms of the assigned Levels of Care.

IAR-DST recommended Level of Care: intake officer vs. clinician rater concordance

For the IAR-DST recommended Level of Care, the inter-rater reliability or concordance between the intake officer and clinician scoring on the tool shows 46.5% agreement, with the clinician rater downgrading 14.1% of the cases and upgrading 39.4% of the cases.

Practitioner-assigned Level of Care: intake officer vs. clinician rater concordance

For the Practitioner-assigned Level of Care, the concordance between the interviewer and clinician rater shows 49.5% agreement with the levels of care, with the clinician rater downgrading 13.1% of the cases and upgrading 37.4% of the cases

Note: Due to technical difficulties for one recording, inter-rater ratings are available for 99 participants.

Rapid Study 2 continued: **Concordance of IAR-DST Domain ratings**

Key findings: The highest agreement at domain level was reached for Domain 2: Harm, followed by Domain 3: Functioning and Domain 4: Impact of co-existing conditions, indicating that strongest agreement was achieved for the IAR-DST Primary assessment domains.

Taking into account the differences between a suitably trained intake officer and a clinician in ratings at the Level of Care, concordance at the Domain level was also examined.

To calculate the level of agreement between intake officers and a clinician rater, we used weighted Cohen's kappa, which assigns different weights to disagreements between the two raters based on their 'distance' on an ordered scale (e.g. none, mild, moderate, severe, etc.). Weighted Kappa is utilised in situations where the categories have a meaningful order, allowing for more nuanced interpretation of agreement.

Weighted Kappa results were interpreted following Cohen (McHugh, 2012) guidelines:

- ≤ 0 no agreement
- 0.01–0.20 none to slight agreement
- 0.21–0.40 fair agreement
- 0.41– 0.60 moderate agreement
- 0.61–0.80 substantial agreement
- 0.81–1.00 almost perfect agreement

Weighted Kappa for each of the eight domains are reported in Table 11. The highest agreement was for Domain 2: Harm at 0.67, followed by Domain 3: Functioning at 0.61, indicating a moderate to substantial agreement. Domain 4: Impact of Co-existing conditions at 0.56 and Domain 5: Service use and response history at 0.54 indicate moderate agreement. The other Domains show fair agreement (0.28-0.40).

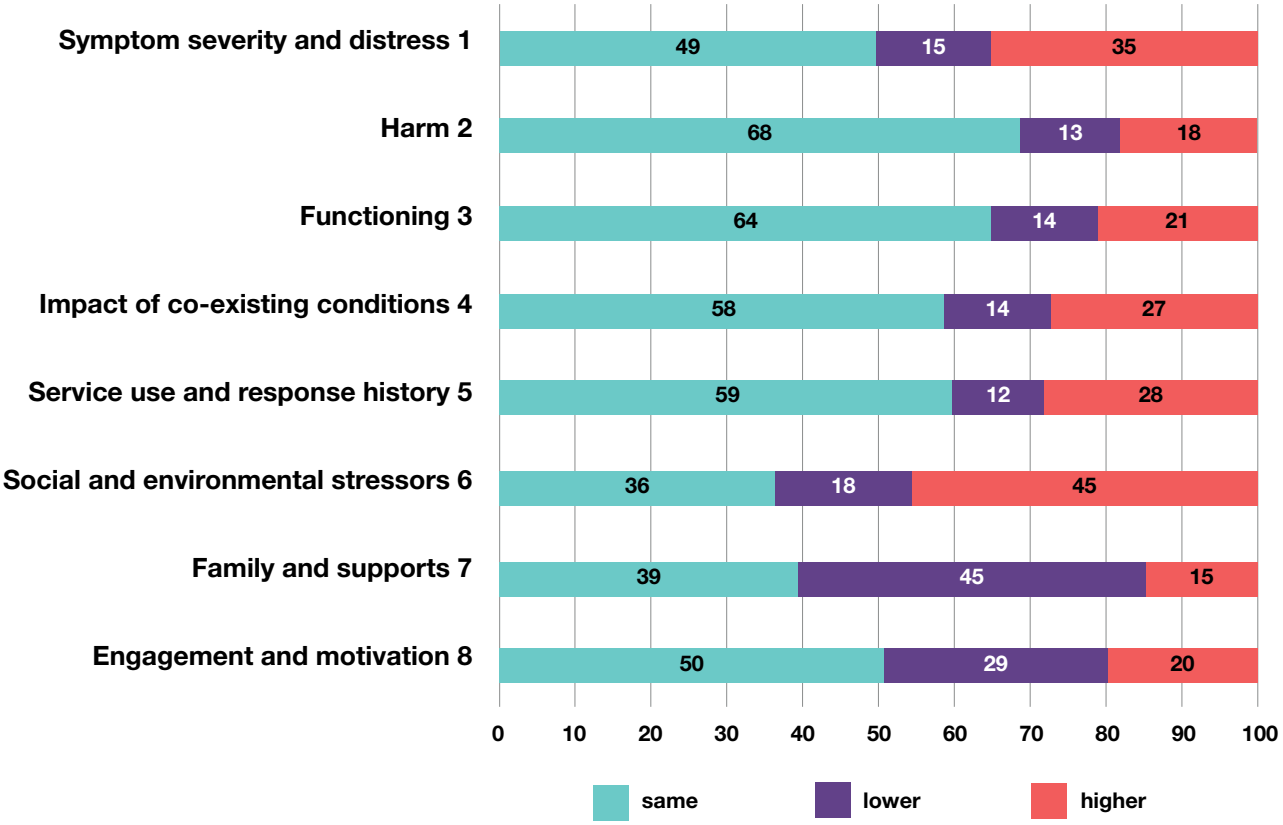
Table 11 . Weighted kappa scores for 8 domains between two raters (n=99).

Domain	Weighted kappa	Significance
1. Symptom severity and distress	0.40	Fair agreement
2. Harm	0.67	Substantial agreement
3. Functioning	0.61	Moderate - Substantial agreement
4. Impact of co-existing conditions	0.56	Moderate agreement
5. Service use and response history	0.54	Moderate agreement
6. Social and environmental stressors	0.31	Fair agreement
7. Family and other supports	0.28	Fair agreement
8. Engagement and motivation	0.30	Fair agreement

Rapid Study 2 continued: **Concordance of IAR-DST Domain Ratings**

Figure 7 shows the inter-rater agreements at the Domain level for the 99 participants, and the instances when the clinician rated a Domain higher or lower than the intake officer. The highest agreement between the two raters was observed for Domain 2: Harm. The clinician rated Domain 6: Social and environmental stressors and Domain 1: Symptom severity and distress considerably higher than the intake officer, whereas Domain 7: Family and other supports was rated considerably lower by the clinician.

Figure 7. Domain-level comparison of ratings between interviewer and clinician rater.



Rapid Study 2 continued: **Standard measures**

Participants completed a comprehensive battery of standardised measures about their mental health, chosen as the closest available match to each of the eight domains. Table 12 presents the measures and their descriptions and the Domains they correspond to.

Table 12. Self-report measures correspondence to domains

Measure	Short name	Focus	Items	Timing	Min	Max	IAR-DST Domain match
DSM-5-TR Self-rated Level 1 Cross-Cutting Symptom Measure	DSM-5-TR	Assess mental health domains	23 (13 domains)	Past 2 weeks	0	Domain specific	1 & 4
WHO Disability Assessment Schedule	WHO-DAS 2.0	Functioning	36 (7 domains)	Past month	8	Domain specific (range 8-40)	3
Alcohol, Smoking, Substance Involvement Screening Test	WHO-ASSIST	Substance use	10 domains	Ever & 3 months	0	Domain specific	4
Non-Suicidal Self-Injury Assessment Tool	B-NSSI-AT	Self-harm	4	Ever	0	15	2
P4 Suicidality screener	P4	Suicidality thoughts	4	Ever	0	N/A	2
Steps to Better Health Questionnaire	SBHQ	Social support	19	Ever	0	8	6 & 7
Sociodemographic questions	N/A	Helpfulness of previous service use	1	Past	1	5	5
Sociodemographic questions	N/A	Motivation to make changes	1	Current	1	5	8

Rapid Study 2 continued: **Standard measures**

In Table 13, we present a summary of the standardised measures and items from the standardised measures that we proposed could correspond to each of the eight domains on the IAR-DST.

Table 13. Description of domains and corresponding measures

Domains	Domain description	Corresponding measure
1. Symptom severity and distress	Symptoms, including internalised (emotional) problems experienced by the person, and externalised behaviours observable by or impacting on others.	DSM-5-TR assessing 13 mental health sub-scales.
2. Harm	Suicidality; intentional, non-suicidal self-harm; impulsive, dangerous, or risky behaviours with the potential for harm to self or others; harm caused by abuse, exploitation, or neglect by others, or by self-neglect.	P4 suicidality screener & B-NSSI-AT self-harm assessing suicidal ideation & self-harm.
3. Functioning	Functional impairment associated with or exacerbated by mental health issues.	WHO-DAS 2.0 providing overview of various aspects of functioning in 7 sub-scales.
4. Impact of co-existing conditions	The extent to which other conditions contribute to (or have the potential to contribute to) increased severity of the mental health issue.	Two items from DSM-5-TR measuring body pain & issues with memory as well as WHO-ASSIST measuring 10 substance (mis)use.
5. Service use and response history	Previous use of services and support focussed on mental health-related assistance.	Single-item sociodemographic question exploring helpfulness of previous service use.
6. Social and environmental stressors	The extent and severity of factors in the person's environment that might contribute to the onset or continuation of the mental health issue.	Three items from the SBHQ measuring financial situation, access to food and suitable housing.
7. Family and other supports	Whether personal supports and social support is present in the person's environment and their potential to contribute to improved mental health.	Two items from the SBHQ measuring social support with family and extended social network.
8. Engagement and motivation	The person's capacity and willingness to engage in or accept assistance.	Single-item sociodemographic question assessing motivation to make changes in mental health.

Domain 1:

Symptom severity and distress

Domain 1 focuses on current symptoms and duration, level of distress, and experiences of mental illness. As Domain 1 is multifaceted, assessing multiple mental health conditions at various levels of severity, we present the concordance with the self-reported *DSM-5-TR Self-Rated Level 1 Cross-Cutting Symptom Measure*, which has 13 sub-scales (depression, anger, mania, anxiety, somatic symptoms, suicidal ideation, psychosis, sleep problems, memory, repetitive thoughts and behaviours, dissociation, personality functioning and substance use) rated on a 5-point Likert scale, yielding 5 levels of severity on the measure that were compared to the intake officer Domain ratings (Table 14).

Table 14. IAR-DST Domain 1 intake officer ratings and self-report measure frequencies

Domain 1	Symptom severity and distress	0 = No problem	1 = Mild symptoms	2 = Moderate symptoms	3 = Severe	4 = Very severe
IAR-DST	Intake officer ratings	2	33	47	17	1
		0 = None	1 = Slight	2 = Mild	3 = Moderate	4 = Severe
DSM-5-TR	Average score across all subscales of DSM-5-TR *	15	38	38	8	1

Table 15. Concordance between Domain 1 & DSM-5-TR total score

IAR-DST 1. Item scale	0 = No problem	1 = Mild Symptoms	2 = Moderate Symptoms	3 = Severe	4 = Very Severe	Total
0-None	1	9	4	1	0	15
1-Mild	1	18	17	2	0	38
2-Moderate	0	5	22	11	0	38
3-Severe	0	1	4	3	0	8
4-Extreme	0	0	0	0	1	1
Total	2	33	47	17	1	100

Results: The proposed match between IAR-DST Domain ratings and severity ratings on the participant self-report measure were the same on 45 occasions, IAR-DST Domain ratings were Higher on 44 occasions and Lower on 11 occasions (Table 15).

* We note Domain 1 requires further disorder-specific analysis.

Domain 2:

Harm

Domain 2 focuses on past or current suicidal ideation, self-harm, self-neglect and risk from others. We averaged across two measures as a proposed match to test concordance for this Domain: the self-reported *P4 Brief Suicidality screener* and self-reported *B-NSSI-AT* which measures non-suicidal self-injury. The average scores across both scales was used for the concordance test (Table 16).

Table 16. IAR-DST Domain 2 interviewer and self-reported measure frequencies

Domain 2	Harm	0 = No concerns about harm	1 = Previous but no current concerns about harm	2 = Some current concerns about harm	3 = Significant current concerns about harm	4 = Very significant current concerns about harm
IAR-DST	Intake officer ratings	38	39	15	7	1
		0 = None	1 = Mild	2 = Moderate	3 = Severe	4 = Extreme
Average score of P4 & B-NSSI-AT	Suicidality & self-harm scales averaged	53	29	17	1	0

Table 17. Concordance between Domain 2 & average scores of P4 & B-NSSI-AT

IAR-DST 2. Item scale	0 = No concerns about harm	1 = Previous but no current concerns about harm	2 = Some current concerns about harm	3 = Significant current concerns about harm	4 = Very significant current concerns about harm	Total
0-None	34	15	4	0	0	53
1-Mild	4	14	7	3	1	29
2-Moderate	0	9	4	4	0	17
3-Severe	0	1	0	0	0	1
4-Extreme	0	0	0	0	0	0
Total	38	39	15	7	1	100

Results: The proposed match between IAR-DST Domain ratings and participant severity ratings on the self-report questionnaires were the Same on 52 occasions, IAR-DST domain ratings were Higher on 34 occasions and Lower on 14 occasions (Table 17). Highest agreement was for Level 0 rating (n=34).

Overall, the intake officer Domain ratings were more severe than the participant self-report measure severity ratings.

Domain 3:

Functioning

Domain 3 focuses on the ability to fulfil usual roles/responsibilities and capacity for self-care. We provide the concordance with the self-reported World Health Organization Disability Assessment Schedule 2.0 (WHO-DAS 2.0) which consists of 36 items and is rated on a 5-point Likert scale. This scale consists of the following subscales: Understanding and communicating, Getting around, Self-care, Getting along with people, Life activities-household, Life activities-school/work and Participation in society (Table 18).

Table 18. IAR-DST Domain 3 interviewer and self-reported measure frequencies

Domain 3	Functioning	0 = No problem	1 =Mild impact	2 = Moderate impact	3 = Severe impact	4 = Very severe to extreme impact
IAR-DST	Intake officer ratings	28	38	24	10	0
		0 = None	1 = Slight	2 = Mild	3 = Moderate	4 = Severe
WHO-DAS	Total scale score *	3	43	39	13	2

Table 19. Concordance between Domain 3 & average scores of WHO-DAS

IAR-DST 3. Item scale	0 = No problem	1 =Mild impact	2 = Moderate impact	3 = Severe impact	4 = Very severe to extreme impact	Total
0-None	2	0	1	0	0	3
1-Mild	19	21	3	0	0	43
2-Moderate	6	15	14	4	0	39
3-Severe	1	2	6	4	0	13
4-Extreme	0	0	0	2	0	2
Total	28	38	24	10	0	100

Results: The proposed match between IAR-DST Domain ratings and participant severity ratings on self-report questionnaires were the Same on 41 occasions, IAR-DST domain ratings were Higher on 8 occasions and Lower on 51 occasions (Table 19).

Overall, the interviewer IAR-DST domain ratings were less severe than the severity ratings according to the participant self-report measure.

* We note Domain 3 requires further analyses separating various aspects of functioning.

Domain 4:

Impact of co-existing conditions

Domain 4 focuses on physical health, intellectual disability, cognitive impairment and substance use/misuse. We tested concordance, by averaging across 2 items from the *DSM-5-TR (Item 9. Unexplained aches and pains'* & '*Item 15. Problems with memory'*) and the total score on the *Alcohol, Smoking, Substance Involvement (WHO-ASSIST)* scale (Table 20).

Table 20. IAR-DST Domain 4 interviewer and self-reported measure frequencies

Domain 4	Impact of co-existing conditions	0 = No problem	1 =Mild impact	2 = Moderate impact	3 = Severe impact	4 = Very severe impact
IAR-DST	Intake officer ratings	28	38	24	10	0
		0 = None (Not at all)	1 =Slight (Rare, less than a day or two)	2 = Mild (Several days)	3 = Moderate (More than half the days)	4 =Severe (Nearly every day)
	Averaged score across DSM-5-TR Items 9 and 15 & WHO-ASSIST *	23	39	30	7	1

Table 21. Concordance between Domain 4 & average scores of three scales

IAR-DST 3. Item scale	0 = No problem	1 =Mild impact	2 = Moderate impact	3 = Severe impact	4 = Very severe impact	Total
0-None	10	8	5	0	0	23
1-Slight	14	17	5	3	0	39
2-Mild	4	12	10	4	0	30
3-Moderate	0	1	4	2	0	7
4-Severe	0	0	0	1	0	1
Total	28	38	24	10	0	100

Results: The proposed match between IAR-DST Domain ratings and participant severity ratings on the self-report questionnaires were the Same on 39 occasions, IAR-DST domain ratings were Higher on 25 occasions and Lower on 36 occasions (Table 21).

Overall, the interviewer IAR-DST domain ratings were less severe than the severity ratings according to the participant self-report measure.

* We note Domain 4 requires further analyses separating various co-existing conditions.

Domain 5:

Service use and response history

Domain 5 focuses on previous treatment, engagement in treatment and response to treatment. We used a single item from the self-reported sociodemographic question (*'How helpful was the treatment you received?'*) rated on a 5-point Likert scale to compare concordance between participant-rated and interviewer-rated scores (Table 22).

Table 22. IAR-DST Domain 5 interviewer and self-reported measure frequencies

Domain 5	Service use and response history	0 = No previous service use	1 = Excellent progress from previous service use	2 = Moderate progress from previous service use	3 = Minor progress from previous service use	4 = Negligible progress from previous service use
IAR-DST	Intake officer ratings	16	25	44	15	0
	Sociodemographic single-rated question	0 = Extremely helpful	1 = Very helpful	2 = Somewhat helpful	3 = Slightly helpful	4 = Not at all helpful
	How helpful was the treatment you received?	20	20	43	11	6

Table 23. Concordance between Domain 5 & average score of single-item question

IAR-DST 5. Helpful?	0 = No previous service use	1 = Excellent progress from previous service use	2 = Moderate progress from previous service use	3 = Minor progress from previous service use	4 = Negligible progress from previous service use	Total
0-No previous use	14	4	2	0	0	20
1-Very / Extremely	0	5	12	3	0	20
2-Somewhat	2	11	22	8	0	43
3-Slightly	0	3	6	2	0	11
4-Not at all	0	2	2	2	0	6
Total	16	25	44	15	0	100

Results: The proposed match between IAR-DST Domain ratings and participant severity ratings on self-report questionnaires were the Same on 43 occasions, IAR-DST scores were Higher on 29 occasions and Lower on 28 occasions (Table 23).

Overall, the interviewer IAR-DST domain ratings were evenly split between more and less severe than the severity ratings according to the participant self-report measure.

Domain 6:

Social and environmental stressors

Domain 6 focuses on life circumstances such as significant transitions, trauma, interpersonal or social difficulties, legal or socioeconomic issues. For concordance we averaged across 3 items from the self-reported Steps to Better Health Questionnaire (SBHQ) ('My financial situation', 'My access to affordable, healthy food', 'My access to suitable housing/accommodation'), which assess various stressors (Table 24).

Table 24. IAR-DST Domain 6 interviewer and self-reported measure frequencies

Domain 6	Social and Environmental Stressors	0 = No problem in this domain	1 = Mildly stressful environment	2 = Moderately stressful environment	3 = Highly stressful environment	4 = Extremely stressful environment
IAR-DST	Intake officer ratings	10	41	37	12	0
	Items from Steps to Better Health Questionnaire	0 = Extremely helpful	1 = Very helpful	2 = Somewhat helpful	3 = Slightly helpful	4 = Not at all helpful
	Items 5, 7, 9 averaged *	26	43	26	5	0

Table 25. Concordance between Domain 6 & average scores of three items

IAR-DST 6.	0 = No problem in this domain	1 = Mildly stressful environment	2 = Moderately stressful environment	3 = Highly stressful environment	4 = Extremely stressful environment	Total
0-None	5	11	8	2	0	26
1-Mild	3	17	16	7	0	43
2-Moderate	2	13	8	3	0	26
3-Severe	0	0	5	0	0	5
4-Extreme	0	0	0	0	0	0
Total	10	41	37	12	0	100

Results: The proposed match between IAR-DST Domain ratings and participant severity ratings on self-report questionnaires were the Same on 30 occasions, IAR-DST scores were Higher on 47 occasions and Lower on 23 occasions (Table 25).

Overall, the interviewer IAR-DST domain ratings were more severe than the severity ratings according to the participant self-report measure.

* We note Domain 6 is complex requiring further analyses separating various stressors

Domain 7:

Family and other supports

Domain 7 focuses on informal supports such as family, friends, peers and supports within community. We chose two items from the self-reported Steps to Better Health Questionnaire (SBHQ), which assessed connection with family and connections with people outside of family. Concordance was calculated by averaging the two items and reversing the scoring to match with the scoring of the IAR-DST Domain 7 (Table 26).

Table 26. IAR-DST Domain 7 interviewer and self-reported measure frequencies

Domain 7	Family & Other supports	0 = Highly supported	1= Well supported	2 = Limited supports	3 = Minimal supports	4 = No supports
IAR-DST	Intake officer ratings	13	40	42	4	1
	Items from Steps to Better Health Questionnaire	0 = Very connected	1 = Mostly connected	2 = Moderately connected	3 = Slightly connected	4 = Not at all connected
SBHQ	Items 1 & 2 averaged	10	32	42	14	2

Table 27. Concordance between Domain 7 & average score of two items

IAR-DST 7.	0 = Highly supported	1= Well supported	2 = Limited supports	3 = Minimal supports	4 = No supports	Total
0-Very connected	3	7	0	0	0	10
1-Mostly connected	3	21	6	2	0	32
2-Moderately connected	7	11	23	1	0	42
3-Slightly connected	0	1	12	1	0	14
4-Not at all connected	0	0	1	0	1	2
Total	13	40	42	4	1	100

Results: The proposed match between IAR-DST Domain ratings and participant severity ratings on self-report questionnaires were the Same on 49 occasions, IAR-DST scores were Higher on 16 occasions and Lower on 35 occasions (Table 27).

Overall, the interviewer IAR-DST domain ratings were less severe than the severity ratings according to the participant self-report measure.

Domain 8:

Engagement and motivation

Domain 8 focuses on ability and capacity to manage condition and motivation to seek help. We used a single item from the self-reported sociodemographic questions ('How motivated are you to work on your current symptoms right now?') rated on a 5-point Likert scale to compare concordance between participant-rated and interviewer-rated scores (Table 28).

Table 28. IAR-DST Domain 8 interviewer and self-reported measure frequencies

Domain 8	Engagement & motivation	0 = Optimal	1 = Positive	2 = Limited	3 = Minimal	4 = Disengaged
IAR-DST	Intake officer ratings	21	64	14	1	0
		0 = Very much	1 = A good amount	2 = Moderately	3 = A small amount	4 = Not at all
SES single-rated question	How motivated are you to work on your current symptoms right now?	24	51	16	8	1

Table 29. Concordance between Domain 8 & average score of two items

IAR-DST 6.	0 = Optimal	1 = Positive	2 = Limited	3 = Minimal	4 = Disengaged	Total
0-Very much	4	18	2	0	0	24
1-A good amount	11	32	8	0	0	51
2-Moderately connected	3	10	3	0	0	16
3-A small amount	3	4	0	1	0	8
4 Not at all	0	0	1	0	0	1
Total	21	64	14	1	0	100

Results: The proposed match between IAR-DST Domain ratings and participant severity ratings on self-report questionnaires were the Same on 40 occasions, IAR-DST scores were Higher on 28 occasions and Lower on 32 occasions (Table 29).

Overall, the interviewer IAR-DST domain ratings were fairly evenly split between more and less severe than the severity ratings according to the participant self-report measure.

Rapid Study 2 continued: **Concordance with standard measures**

Concordance summaries between interviewer and self-reported scores

In addition to observing the concordance between the interviewer-generated IAR-DST recommended level and self-reported scores, we conducted weighted kappa analyses to observe **inter-rater reliability between the two scores** (Table 30).

Table 30. Concordance summaries for each domain between IAR-DST and self-reported measure scores

IAR-DST Domains	Same score between interviewer and self-reported measures	Interviewer rating higher compared to self-reported measures	Interviewer rating lower compared to self-reported measures	Weighted kappa	Significance
1. Symptom severity & distress	45	44	11	0.31	Fair agreement
2. Harm	52	34	14	0.39	Fair agreement
3. Functioning	41	8	51	0.33	Fair agreement
4. Impact of coexisting conditions	39	25	36	0.21	Slight to Fair agreement
5. Service use and response history	43	29	28	0.34	Fair agreement
6. Social & environmental stressors	30	47	23	0.05	None to Slight agreement
7. Family & other supports	49	16	35	0.34	Fair agreement
8. Engagement & motivation	40	28	32	0.03	None to Slight agreement

We observed that for Primary assessment domains (Domains 1, 2, 3) and for Domains 5 and 7, there was fair inter-rater reliability between self-reported symptom scores and interviewer-recommended scores on the IAR-DST.

Key findings: Inter-rater reliability tests showed fair agreement between the self-rated standardised measures and interviewer-rated IAR-DST scores for 3 of the 4 primary domains.

Key findings: Participants felt comfortable having their mental health needs assessed, they felt heard, respected and listened to, and found that the questions asked pertaining to the 8 domains covered all areas of their support needs.

All participants who completed the IAR-DST with an interviewer were asked to complete a post-interview survey, which consisted of four Likert-type questions and one open-ended question where participants could provide additional information about their experiences of completing the mental health needs assessment with the interviewer (Table 31).

The majority of participants reported they felt comfortable having their mental health needs assessed (n=99) and felt they were listened to in all aspects of their support needs (n=98). Additionally, the majority of participants felt the questions asked during IAR-DST assessment covered all areas of their support needs (n=91) and the overall process made them feel they could get appropriate support for their needs (n=84).

Table 31. Post-IAR-DST-interview survey responses (n=100)

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1. You felt comfortable in your mental health needs assessment.	65	34	1	0	0
2. You were listened to in all aspects of your support needs (incl. physical & mental health, housing, employment, etc).	70	28	0	2	0
3. The questions that were asked in your mental health needs assessment covered all areas of your support needs.	50	41	7	1	1
4. The process made you feel that you could get appropriate support for your needs.	44	40	12	4	0

Participant feedback after interviews

Open-ended feedback revealed that participants felt comfortable sharing their experiences, that they felt they were listened to, and that their needs were understood. Participants specifically appreciated the clarity of the questions and the examples given, which made the difficult subject matter more approachable. Although most participants reported that they felt the interview was long enough to cover their needs respectfully but without being exhausting, some respondents wanted a longer, more in-depth assessment.

Rapid Study 3

Clinician survey

Key findings: Responses from health professionals who use the IAR-DST were mixed. Overall, the feedback from those who use the IAR-DST to make referrals was more negative, and the feedback from those who use the tool to receive referrals was more positive.

Results in context: The feedback from those who make referrals was mostly from GPs who reported they were reasonably or very experienced in mental health assessments.

Rapid study 3 was an anonymous online survey for clinicians who use the IAR-DST to receive or make referrals, which provides further evidence of face validity and the general utility of the tool by exploring the experiences of clinicians of the usefulness of IAR-DST in their clinical practice.

A total of 94 clinicians took part in the survey: GPs (59%), PHN commissioned service providers (12%), Medicare Mental Health Centre staff members (10%), phone line staff (9%) and headspace staff members (10%). Almost all staff were reasonably (48%) or very experienced (47%) in mental health. The majority of respondents made referrals (86%) and a third received referrals (33%).

Clinicians making referrals

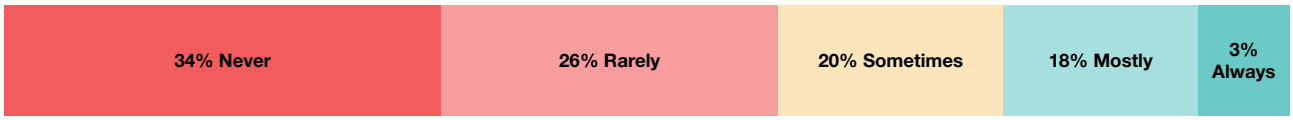
Of those clinicians who made referrals, nearly half of the respondents (n=39, 49%) believed that the IAR-DST mostly or always produced the right Level of Care. However, 60% (n=48) said that the IAR-DST never, or only rarely, helped to create a common language about levels of care across service types, and over 60% (n=49) stated that the IAR-DST never, or rarely, helped them to decide about which service to refer someone to.

Figure 8. Clinicians making referrals (n=80)

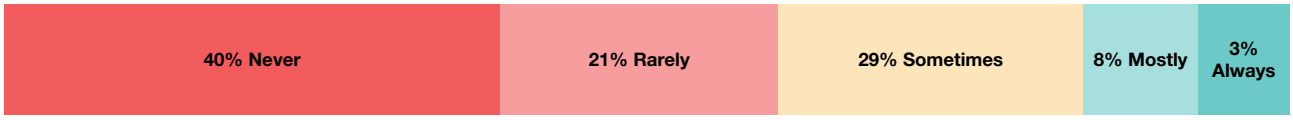
Does the IAR-DST produce the right level in your clinical opinion?



Has the IAR-DST helped to create a common language about levels of care across service types?



Does the IAR·DST help you to make a decision about the type of service to refer someone to?



Clinicians making referrals continued

The majority of respondents did not think the IAR-DST is a useful tool that enhances their practice (63%), did not think the tool is time-efficient in their practice (69%), did not think use of the tool enabled access to service pathways that otherwise would not be accessible (64%), and nearly half thought that the IAR-DST had further complicated their referral processes (49%).

Clinicians receiving referrals

Of the respondents who received referrals, nearly half (48%) thought that the IAR-DST had mostly recommended the right Level of Care, that the referrals were mostly appropriate for the type of care that they could deliver, and that patients referred an outcome from the IAR-DST could mostly achieve good outcomes from their care. In their opinion, patients received the right care from the system either some of the time (48%) or most of the time (39%) (Figure 9).

Figure 9. Clinicians receiving referrals (n=31)

When you receive a referral that includes an IAR-DST rating, has the tool recommended the right Level of care?



Are the referrals that you receive using the IAR·DST appropriate for the type of care that you can deliver?



Do patients who are referred using the IAR·DST receive the right care from the system?



Do patients who are referred using the IAR-DST achieve good outcomes from their care?



Clinicians receiving referrals continued

About half of respondents considered the IAR-DST to have improved referral processes (52%), and 42% reported that it had improved the quality of information received in a referral. Nearly all respondents reported receiving sufficient training in the use of the IAR-DST (94%) (Table 32).

Table 32. Clinicians receiving referrals (n=31)

Is the IAR-DST an improvement on standard referral practices?			***Has the IAR-DST improved the quality of information you receive in a referral?***			***Have you had training in the IAR-DST sufficient to understand its framework and how to use?***		
It has improved the referral process	**16**	**52%**	It has improved the quality of information in referrals	**13**	**42%**	No training	0	0
It has not improved the referral process	9	29%	About the same	11	36%	Some training but insufficient	2	6.5%
It has made referral processes worse	6	19%	It has decreased the quality of information I receive in referrals	7	23%	Sufficient training	**29**	**93.5**

However, clinicians who receive referrals also reported the IAR-DST was mostly used to derive the level of care specified to access a service (84%).

Table 33. Clinicians receiving referrals (n=31)

Have you observed the IAR-DST being used in any of the following ways (tick all that apply):		
To derive the care level specified to access a service	26	84%
As a proxy for outcome	6	19%
As a reassessment tool to move people to the next step in care	10	32%
Instead of other more appropriate assessments being completed	11	36%

Finally, most clinicians who receive referrals reported that they did not have restrictions on their practice relating to IAR-DST levels of care (62%), of those who did report restrictions, (e.g. that they can only treat people assigned certain levels of care as determined by the IAR-DST), a proportion stated that this was causing both clinical and administrative problems (23%).

References

Australian Government, Department of Health and Aged Care. (17 March 2022). National Initial Assessment and Referral (IAR) for Mental Healthcare Project. Expert Advisory Group – Terms of Reference. Available from: https://www.acrrm.org.au/docs/default-source/all-files/national-iar-eag-terms-of-reference.pdf [EAG]

Australian Government, Department of Health and Aged Care. (May 2024). Initial Assessment and Referral (IAR) Guidance for Mental Health. Part A – General Guidance (Version 2.0). Available from: https://emphn.org.au/wp-content/uploads/2024/07/IAR-Guidance-for-Mental-Health-Part-A-General-Guidance-v2.0.pdf [EAG]

Australian Psychological Society (APS). (20 December 2021). Letter in response to Consultation Draft – National Initial Assessment and Referral for Mental Healthcare – Child and Adolescent. Available from: https://psychology.org.au/getmedia/0cffde66-cd6c-4368-ac13-7b982e900144/20211220-aps-submission-national-initial-assessment-and-referral-for-mental-healthcare-for-children-and-adolescents_web.pdf [EAG]

Australian Psychological Society. (2018). Assessment, triage, and referral processes in stepped care mental health systems: A literature review. Report prepared for the Australian Government Department of Health, Canberra (report approved in 2018 & updated on 2. August 2021)

Head to Health (14 March 2024). Head to Health Phone Service, Insights Report (July 2022 to December 2023)

McHugh M. L. (2012). Interrater reliability: the kappa statistic. Biochemia medica, 22(3), 276–282.

Mokkink L.B., de Vet H.C.W., Prinsen C.A.C., Patrick D.L., Alonso J., Bouter L.M., Terwee C.B. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Quality of Life Research. 2018;27(5):1171–1179. doi: 10.1007/s11136-017-1765-4.

Nous Group [commissioned by Australian Government, Department of Health and Aged Care]. (19 September 2023). IAR Review Project, Interim Insights Report.

Prinsen C.A.C., Mokkink L.B., Bouter L.M., Alonso J., Patrick D.L., de Vet J.C.W., Terwee C.B. COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research.

2018;27(5):1147–1157. doi: 10.1007/s11136-018-1798-3.

Terwee C.B., Prinsen C.A.C., Chiarotto A., Westerman M.J., Patrick D.L.…Mokkink L.B. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. Quality of Life Research. 2018;27(5):1159–1170. doi: 10.1007/s11136-018-1829-0.

University of Melbourne, Victoria. (25 March 2021). Report for the Implementation Review of the National Initial Assessment and Referral Guidance for the Department of Health.

INSIDE OUT
Institute for Eating Disorders