PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Informant-based assessment instruments for dementia and their measurement properties in persons with intellectual disability: a systematic review protocol
AUTHORS	Zeilinger, Elisabeth; Komenda, Sophie; Zrnic, Irina; Franken, Fabian; Woditschka, Katharina

VERSION 1 – REVIEW

REVIEWER	Wieneke Mokkink Amsterdam UMC, location VUmc, dep Epidemiology and data science, Amsterdam, the Netherlands I'm one of the developers of the COSMIN methodology that will be used in this review.
REVIEW RETURNED	30-Jun-2020

GENERAL COMMENTS	Manuscript BMJopen-2020-040920 describes the protocol of a systematic review on instruments to measure dementia in people with intellectual disabilities. In general, the review will be conducted using up-to-date methods. I have some comments and questions for clarification. The authors use the term Informant-based assessment instrument. Walton et al (DOI: 10.1016/j.jval.2015.08.006) defines different types of instruments, using the term Observer-reported outcome measures (ObsROM). Is the term 'informant-based assessment instrument' being used as a synonym, meaning that the proxy completes the measure? As the construct of interest is dementia, I was confused, as to me it seems to be a diagnosis, and it is not up to a non-clinical proxy to set a diagnosis. Perhaps it is not used as a diagnostic tool, but rather an outcome measure to assess e.g. severity of dementia (e.g. various symptoms and consequences on e.g. physical or social function)? If it is not used as a synonym for ObsROM, do the authors refer to diagnostic instruments, completed by a clinician, but using proxy input? Why is the first aim added? This aim has consequences for the search strategy, and (probably) the amount of work to be done (see also my later comment). An elaboration on the specific purpose would be interesting to read, e.g. a content analysis (scoping review) will be done to be used as input for the development of a new instrument. The aims 2-3 refer to what is being done, rather than the 'ultimate' goal of such a review, which is to do a recommendation of the most suitable measurement instruments to measure 'dementia' in people with intellectual disabilities. I recommend the authors to do

	a recommendation on the most useful instrument(s) in the end, based on the whole overview they have made. The search strategy consists of two phases. First, all 'informant-based' instruments measuring dementia are searched, also when no studies on measurement properties will exist. Second, all studies on development and measurement properties of this set of instruments will be searched. In my experience, this is more work than to combine it in one search, using the COSMIN search filter for finding studies on measurement properties. Both strategies are fine, however, it should also be doable. Why is the search filter not used? Do the authors expect to find important instruments that are being used, but for which never formally the measurement properties were assessed? Or, because of translation and availability issues of the filter to other databases than PubMed and EMBASE? Was a clinical librarian involved in the development of the search string? (I strongly recommend this!) Page 7 line 150-152: '… various terms for the (1) output of interest, (2) measure of interest, and (3) the specified population.' The first two terms were not clear to me. In Table 1 'output' is operationalized using both terms for the type of instrument (e.g. questionnaire), and the purpose of use (e.g. diagnostic and screening), while 'measure of interest' refers to the outcome of interest (i.e. dementia). Why are words like 'outcome' and 'evaluation' not used in the search string (column 'output')? Why is the focus of the instrument only on diagnostic or screening instrument? COSMIN is for outcome measurement instruments, while QUADAS or QAREL is for diagnostic use of instruments. Page 10: the COSMIN checklists [26–28] and the CAPs-IDD [22] will be combined. How will they be combined? And to what kind of quality ratings do the authors refer on page 10 line 215 'quality ratings according to CAPs-IDD'? Minor issues: Page 5, line 120/121: 'existing body of research'. To what kind of research do you refer, also to all studies is which any of the instruments found was used? Is this feasible? Page 8 line 163-164: '(2) include at least one informant-based instrument (development or evaluation) for the assessment of dementia'. What is meant by '(development and evaluation)'? Page 8 line 164-165: '(3) this instrument has to be especially developed or adapted for persons with ID'. Can you know this? If it was developed for another patient population, and not adapted, but used, and evaluated in patients with ID, will it be excluded? Page 8 line 170-171: 'Once we have identified the instruments, we will conduct a search by citation strategy using the initial publications of each instrument as a reference point'. what is a citation strategy? How will you find all initial publications? You may not found the original publication from the first search, e.g. as it was developed for another population. Page 9 line 190: 'We will re-run the search before the final analyses'. Do the authors refer to both searches, or only to the second search?

	Page 9 line 201: 'We will include all studies, irrespective of their design, but apply the COSMIN quality rating'. To which rating do the authors refer, the Risk of Bias ratings, or the ratings for sufficient quality of the instruments? Page 10 line 209: '… on study level, on single outcome level, and on an aggregated outcome level, …'. What is meant by 'single outcome level' and 'aggregated level'? Within the cosmin methodology you assess the risk of bias only on study level. The quality of the instrument is assessed at study level (i.e. per instrument per measurement property per study), and on the summarized result (i.e. per measurement property per instrument.) Page 10 line 216: 'interrater agreement will be determined using Cohen's Kappa'. Kappa refers to reliability, not to agreement. Percentage specific agreement is a better measure (see the COSMIN standards). Page 10 line 218: 'the quality rating of the studies will go into the final appraisal of the quality of available evaluation data for each instrument.' The quality ratings of the studies refers to the Risk of Bias rating. what is meant by 'final appraisal of the quality of available evaluation data'? does this refer to CAPsIDD ratings? Page 10 Line 221: 'published by the author(s) of the respective instrument'. Do the authors mean the developers of the respective instruments?

REVIEWER	Rawan AlHeresh MGH Institute of Health Professions United States
REVIEW RETURNED	07-Jul-2020

GENERAL COMMENTS	This manuscript entitled Informant-based assessment instruments for dementia and their measurement properties in persons with intellectual disability: a systematic review protocol describes the intended execution to take place between May and August of 2020. This is a very much needed area of evidence, and the findings of the proposed protocol would eventually assist clinicians and researchers in identifying sound patient-reported measures of intellectual functioning for people with dementia. The following would contribute to strengthening the methods and the overall premise of the manuscript: - Since the main target of inquiry is very patient informed I encourage the authors to look into and integrate the International Classification of Functioning Disability, and Health framework; as functioning (ICF) (including intellectual) would give a foundation to how it should be assessed from a capacity perspective as opposed to focusing on impairment (like the ICD-11, DSM-5…etc.) - Lines 87-88, if the presentation dementia is different for people with ID, then new measures would be warranted as they other would not applicable? - Line 95 "In contrast to direct tests of cognitive functioning," this is confusing. Cognitive functioning can be attained too using patient/parent-reported measures in various fields such as rehabilitation. Please revise to reflect this, and make the case of needing this review without eluding to this point.

	- Line 29 and 30, not very clear- some PROs are applicable (performance-based functional assessments, for instance), IRT based ones..etc. - Lines 110-114: Not convincing, since there have been other reviews published in the past as you mention in this area. Explicitly say what is the “very inclusive” approach is, and how it is superior? How will it add to the existing body of knowledge? What does the evaluation method you used add here? - I think you need to describe the discrepancy in the literature on how and why you think there are variations (if they exist) in measuring/detecting dementia. This would serve as a reason to why you are conducting this systematic review. - Objective one seems unnecessary since it has been done in the past; what will you differently? - Introduction literature is outdated, include studies less than five years old to support your need in the introduction. Methods and analysis: - Lines 131-133 can be merged into one sentence. - Table 1: Add the term “Client-report or person-centered or patient-report” see PROMIS methodology for identifying search constructs: https://www.healthmeasures.net/explore-measurement-systems/promis - For Medline, include MeSH heading search to be “more inclusive.” - Lines 167 and 168, not clear why you can not include scales for screening and stratify them for your intended study? (You include the screen in your search strategy?) - Include specific measurement properties in the search strategy, such as “validity, reliability, responsiveness…ect.” - Not clear what the difference is between search one and two, and not clear what happens between the two, will all studies go through the two inclusionary criteria? More clarification is needed. No information on who and how many reviewers there will be per stage, all seems merged. - Line 193: There needs to be a method set a-priori for disagreement dissonances, i.e., a third reviewer? - Data extraction: As per the COSMIN guidelines, you need to include the country and the cross-cultural adaptability of the instruments identified in the data extraction. - Line 206: what does important mean? Avoid using terms like this as it is a matter of relevance. - Line 227, “data pooling not possible.” Look into this method of qualitative synthesis of measurement properties: https://www.jclinepi.com/article/S0895-4356(06)00174-0/fulltext - Line 246: Zotero can also be used to manage duplication of titles/studies from multiple databases. Regarding the limitations section, study limitations are listed without a discussion on the impact/resolutions of these limitations may have on the outcome of the research. - Overall: A few spelling mistakes, please check the language.

REVIEWER	Laura Hughes University of Sussex, United Kingdom
REVIEW RETURNED	26-Aug-2020

GENERAL COMMENTS	This is a well-written protocol paper for a systematic review of informant based dementia assessment instruments for people with intellectual disabilities. The content of the review will likely be of value to those researching and working clinically with people with ID and dementia. The introduction clearly lays out the rationale for the review, including discussion for conducting an up-to-date more focussed review on ID and dementia specific instruments. One point to consider is to move or remove the paragraph about COSMIN and CAPs-IDD from the introduction as it already fits well in the methods section and where it is currently situated detracts from the rest of the introduction as it is between the rationale and objectives. Methods, analysis and data extraction are well structured and detailed. I am not sure if you need the statement about PPI within the text, the statement about lack of resources does not sit right, it might be more appropriate to state that PPI was not carried out due to the scope of a systematic review of the literature.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Wieneke Mokkink
Institution and Country: Amsterdam UMC, location VUmc, dep Epidemiology and data science, Amsterdam, the Netherlands
Competing interests: I'm one of the developers of the COSMIN methodology that will be used in this review

Please leave your comments for the authors below
Manuscript BMJopen-2020-040920 describes the protocol of a systematic review on instruments to measure dementia in people with intellectual disabilities. In general, the review will be conducted using up-to-date methods. I have some comments and questions for clarification.

The authors use the term Informant-based assessment instrument. Walton et al (DOI: 10.1016/j.jval.2015.08.006) defines different types of instruments, using the term Observer-reported outcome measures (ObsROM). Is the term 'informant-based assessment instrument' being used as a synonym, meaning that the proxy completes the measure? As the construct of interest is dementia, I was confused, as to me it seems to be a diagnosis, and it is not up to a non-clinical proxy to set a diagnosis. Perhaps it is not used as a diagnostic tool, but rather an outcome measure to assess e.g. severity of dementia (e.g. various symptoms and consequences on e.g. physical or social function)? If it is not used as a synonym for ObsROM, do the authors refer to diagnostic instruments, completed by a clinician, but using proxy input?

> *Thank you for these thoughtful questions! If the definitions established be Walton et al. are considered, our review includes observer-reported outcome measures (ObsROM), as well as clinician reported outcome measures (ClinROM). In the field of dementia assessment in persons with ID, a distinction is commonly drawn between three kinds of instruments: direct cognitive tests, informant-based assessment instruments (which are also called observer rated scales), and medical test (like fMRI, gene-markers, etc.). With the present review we*

focus on informant-based instruments, only. Informants can be proxies without professional clinical training as well as clinical professionals. In the results we will depict (among other details) for each instrument who the informant should be, or if special training is needed.

In the revised version we clarified this aspect in the introduction and refer to the types of instruments established by Walton et al..

Why is the first aim added? This aim has consequences for the search strategy, and (probably) the amount of work to be done (see also my later comment). An elaboration on the specific purpose would be interesting to read, e.g. a content analysis (scoping review) will be done to be used as input for the development of a new instrument.

Our first aim is to identify informant-based instruments suitable for the assessment of dementia in persons with ID. The rationale of this aim is to provide an overview of all available informant-based dementia assessment instruments for persons with ID. We believe that knowledge about the availability and characteristics of developed/published scales is of great value for both practitioners and researchers. We want to provide such an overview. We added this explanation in the introduction of the revised manuscript.

The aims 2-3 refer to what is being done, rather than the 'ultimate' goal of such a review, which is to do a recommendation of the most suitable measurement instruments to measure 'dementia' in people with intellectual disabilities. I recommend the authors to do a recommendation on the most useful instrument(s) in the end, based on the whole overview they have made.

We agree with the reviewer that our aims resembled a description of the procedure. We changed the wording of these aims to better reflect the true goals of our work. We included the aim to provide a recommendation on the most useful instruments.

The search strategy consists of two phases. First, all 'informant-based' instruments measuring dementia are searched, also when no studies on measurement properties will exist. Second, all studies on development and measurement properties of this set of instruments will be searched. In my experience, this is more work than to combine it in one search, using the COSMIN search filter for finding studies on measurement properties. Both strategies are fine, however, it should also be doable. Why is the search filter not used? Do the authors expect to find important instruments that are being used, but for which never formally the measurement properties were assessed? Or, because of translation and availability issues of the filter to other databases than PubMed and EMBASE?

We are aware that our search strategy amounts to a considerable workload for the review team. However, we decided to take on this workload to be as inclusive as possible with our search strategy. As the first author has expert knowledge in the field of dementia assessment in persons with ID, she is aware that there are numerous instruments without any proper evaluation of measurement properties. The use of search filters would lead to an omission of these instruments. It is one goal of the present review to depict the amount of research done for evaluating dementia assessment instruments for persons with ID. Unfortunately, we expect the outcome of this review to point to huge gaps in (desperately needed) evaluation studies. To be able to point this out, we need to include all available informant-based dementia instruments for persons with ID, irrespective of the existence of evaluation studies.

Was a clinical librarian involved in the development of the search string? (I strongly recommend this!)

> *As this project has no funding, we could neither include a clinical librarian, nor persons with ID or the public. However, among the review team there are clinical psychologists and neuropsychologists with expert knowledge in the field of dementia and ID, including dementia assessment, as well as experts on systematic reviews and psychometrics.*

Page 7 line 150-152: '… various terms for the (1) output of interest, (2) measure of interest, and (3) the specified population.' The first two terms were not clear to me. In Table 1 'output' is operationalized using both terms for the type of instrument (e.g. questionnaire), and the purpose of use (e.g. diagnostic and screening), while 'measure of interest' refers to the outcome of interest (i.e. dementia).

> *We agree with the reviewer that the headings of the table were misleading. We changed "measure" to "construct" (which is "dementia" in our review). However, we believe that the heading "output" is adequate, as it refers to the output of interest in this study, which are assessment instruments. The synonyms listed for "assessment instruments" include the type of instruments and the purpose of use. For example, "diagnostic instrument" or "screening instrument" are different types of instruments, although the purpose can also be derived from the names. We therefore group the type and purpose of instruments under the same category (output).*

Why are words like 'outcome' and 'evaluation' not used in the search string (column 'output')? Why is the focus of the instrument only on diagnostic or screening instrument? COSMIN is for outcome measurement instruments, while QUADAS or QAREL is for diagnostic use of instruments.

> *As described above, we want to include all instruments, even those without any evaluation (and unfortunately, there are some of those). Therefore, we do not include words like "evaluation" in the first search. Evaluation studies of instrument will be found via the second search.*
>
> *The target of our review are assessment instruments for dementia in persons with ID, completed by informants. These are mostly screening instruments, which are used to track dementia-related changes in an individual or to gather information that can lead to a referral for dementia diagnosis. These instruments are not diagnostic instruments and they do not lead to a diagnosis. A diagnosis of dementia can only be made by a professional (clinical psychologist, psychiatrist) and a clinical interview. QUADAS and QAREL are not adequate for the types of instruments in our study, whereas COSMIN fits perfectly.*

Page 10: the COSMIN checklists [26–28] and the CAPs-IDD [22] will be combined. How will they be combined? And to what kind of quality ratings do the authors refer on page 10 line 215 'quality ratings according to CAPs-IDD'?

> *We will use the COSMIN methodology and accompany it with PART I of the CAPs-IDD. We do not use PART II of the CAPs-IDD, as this part comprises psychometric properties, which are more comprehensively covered by the COSMIN ratings. However, PART I of the CAPs-*

IDD provides descriptive features especially designed for measurement instruments on psychiatric disorders for persons with ID (e.g. who is the informant, for what level of disability is the instrument meant to be used, etc.). These descriptive features will be especially useful for practitioners (and researchers) when providing an overview of all available instruments (aim 1 of this study).

In the revised manuscript we explained the combination of COSMIN and CAPs-IDD in more detail. We also rephrased the sentence including "…quality ratings according to CAPs-IDD…", as it was misleading.

Minor issues:

Page 5, line 120/121: 'existing body of research'. To what kind of research do you refer, also to all studies is which any of the instruments found was used? Is this feasible?

We agree with the reviewer that this sentence is poorly phrased and our study can be sufficiently explained by the sentences following this one. We amended the respective paragraph accordingly.

Page 8 line 163-164: '(2) include at least one informant-based instrument (development or evaluation) for the assessment of dementia'. What is meant by '(development and evaluation)'?

With this sentence, we wanted to express that we only include studies describing an evaluation or a development of an instrument. We rephrased it in the revised manuscript to make this clearer.

Page 8 line 164-165: '(3) this instrument has to be especially developed or adapted for persons with ID'. Can you know this? If it was developed for another patient population, and not adapted, but used, and evaluated in patients with ID, will it be excluded?

If an instrument was evaluated in persons with ID, but was not meant to be used with this special population, we will not include this instrument in our review. It is not best-practice to use dementia instruments in persons with ID, that were not particularly designed or adapted for this population. Commonly, this information about an instrument is described in the respective papers. If it is not described, we will seek further information on the respective instrument to determine if it was especially developed or adapted for persons with ID.

Page 8 line 170-171: 'Once we have identified the instruments, we will conduct a search by citation strategy using the initial publications of each instrument as a reference point'. what is a citation strategy? How will you find all initial publications? You may not found the original publication from the first search, e.g. as it was developed for another population.

A "search by citation strategy" is a method of citation searching. We use the initial publication as a starting point and search through all records that have cited this initial publication. The first search is meant to provide an inventory of instruments. We use the initial publication of these instruments for the second search. To give an example: If we find a paper evaluating the instrument XYZ in search one, we determine the initial publication of instrument XYZ (irrespective if it was part of search 1) and use this initial publication as starting point for the second search. The initial publication can be more than one. E.g., if there is a published

manual and a paper describing instrument development, we use both as starting point for the second search.

Page 9 line 190: 'We will re-run the search before the final analyses'. Do the authors refer to both searches, or only to the second search?

We will re-run both searches. We pointed this out in the revised manuscript.

Page 9 line 201: 'We will include all studies, irrespective of their design, but apply the COSMIN quality rating'. To which rating do the authors refer, the Risk of Bias ratings, or the ratings for sufficient quality of the instruments?

With this sentence we referred to the Risk of Bias ratings, as it captures the methodological quality of studies. However, examining this sentence on detail, we think it is not necessary, even confusing, to mention this here. We go into detail with our coding procedure in the "Risk of bias and quality assessment" section. Therefore, we deleted the confusing part of the sentence, and it now reads: "We will include all studies, irrespective of their design."

Page 10 line 209: '… on study level, on single outcome level, and on an aggregated outcome level, …'. What is meant by 'single outcome level' and 'aggregated level'? Within the cosmin methodology you assess the risk of bias only on study level. The quality of the instrument is assessed at study level (i.e. per instrument per measurement property per study), and on the summarized result (i.e. per measurement property per instrument.)

We agree that our description was not clear. We revised it.

Page 10 line 216: 'interrater agreement will be determined using Cohen's Kappa'. Kappa refers to reliability, not to agreement. Percentage specific agreement is a better measure (see the COSMIN standards).

Thank you for this comment. We changed our plan accordingly, and will calculate percentage agreement to determine interrater agreement.

Page 10 line 218: 'the quality rating of the studies will go into the final appraisal of the quality of available evaluation data for each instrument.' The quality ratings of the studies refers to the Risk of Bias rating. what is meant by 'final appraisal of the quality of available evaluation data'? does this refer to CAPsIDD ratings?

Again, we apologize for the unclarity, and admit that the wording was not adequate. We wanted to express that the ratings of each study (using the Risk of Bias rating) will be combined to an overall appraisal of instruments and of available evaluation studies. However, this is just a repetition of our study aim. We deleted this sentence, as it contains no new information.

Page 10 Line 221: 'published by the author(s) of the respective instrument'. Do the authors mean the developers of the respective instruments?

Yes. We changed the wording of this sentence to be clearer.

Reviewer: 2
Reviewer Name: Rawan AlHeresh
Institution and Country:
MGH Institute of Health Professions
United States
Competing interests: None declared

Please leave your comments for the authors below
This manuscript entitled Informant-based assessment instruments for dementia and their measurement properties in persons with intellectual disability: a systematic review protocol describes the intended execution to take place between May and August of 2020. This is a very much needed area of evidence, and the findings of the proposed protocol would eventually assist clinicians and researchers in identifying sound patient-reported measures of intellectual functioning for people with dementia. The following would contribute to strengthening the methods and the overall premise of the manuscript:

Thank you very much for your kind evaluation. We are afraid that the purpose of our study was not clearly communicated. We are not looking at patient-reported measures, but on informant-based measures. Furthermore, those measures should not assess intellectual functioning in persons with dementia. They should assess /screen for dementia in persons with an intellectual disability (ID). We are very sorry that we have not made this clear enough in the paper. In the revised version, we added some additional clarifications in the introduction.

- Since the main target of inquiry is very patient informed I encourage the authors to look into and integrate the International Classification of Functioning Disability, and Health framework; as functioning (ICF) (including intellectual) would give a foundation to how it should be assessed from a capacity perspective as opposed to focusing on impairment (like the ICD-11, DSM-5…etc.)

The target is not patient informed, as explained above.

- Lines 87-88, if the presentation dementia is different for people with ID, then new measures would be warranted as they other would not applicable?

This is true, and is one of the reasons for this review. We are discussing this aspect at several point in the manuscript, including the following statement in the introduction: "Well-evaluated assessment and screening instruments for the general population, such as the frequently used Mini-Mental State Examination (MMSE)[11] are not suitable for persons with ID due to their pre-existing disabilities.[12,13]"

- Line 95 "In contrast to direct tests of cognitive functioning," this is confusing. Cognitive functioning can be attained too using patient/parent-reported measures in various fields such as rehabilitation. Please revise to reflect this, and make the case of needing this review without eluding to this point.

We agree with the reviewer. Again, we are afraid we did not make ourselves clear enough. We do include measures on cognitive functioning, but only if they are informant-based. We do not included measures that include tests of cognitive functioning like the MMSE, CERAD, etc. We depicted this in the aim of the review in in the inclusion criteria.

- Line 29 and 30, not very clear- some PROs are applicable (performance-based functional assessments, for instance), IRT based ones..etc.

We are afraid that this comment is linked to our failure to communicate the target population clearly enough. For persons with ID, most dementia assessment methods used for the general population (persons without ID), are not applicable.

- Lines 110-114: Not convincing, since there have been other reviews published in the past as you mention in this area. Explicitly say what is the "very inclusive" approach is, and how it is superior? How will it add to the existing body of knowledge? What does the evaluation method you used add here?

So far, no review on dementia instruments for persons with ID included an evaluation system like COSMIN and the CAPs-IDD, and no review used such an inclusive search-strategy as ours. These aspects are already explained in the manuscript.

- I think you need to describe the discrepancy in the literature on how and why you think there are variations (if they exist) in measuring/detecting dementia. This would serve as a reason to why you are conducting this systematic review.

This is described in the introduction. However, we made this even clearer now by adding some more information.

- Objective one seems unnecessary since it has been done in the past; what will you differently?

*There is one review listing all assessment instruments for dementia in persons with ID (*https://doi.org/10.1016/j.ridd.2013.08.013*). It was published 2013, 7 years ago. Our first aim (as already described in the paper) is to update this review and provide an inventory of available assessment instruments for persons with ID. This will be of great value to practitioners and researchers in this field.*

- Introduction literature is outdated, include studies less than five years old to support your need in the introduction.

We have used and cited literature that is relevant to our study and important in our field of research. At this point, we would also like to indicate that we have included several studies published in the last five years (e.g.: doi:10.1002/gps.5258; doi:10.1111/jar.12441; doi:10.2174/1567205012666150921095724)

Methods and analysis:

- Lines 131-133 can be merged into one sentence.

We merged this into one sentence.

- Table 1: Add the term “Client-report or person-centered or patient-report” see PROMIS methodology for identifying search constructs: https://www.healthmeasures.net/explore-measurement-systems/promis

We do not include patient-centred, nor patient-reported measures, hence adding this term would be misleading However, thank you for this interesting reference.

- For Medline, include MeSH heading search to be “more inclusive.”

We want our searches to be comparable over all databases. Using MeSH heading in Medline makes the search more restrictive. This might save some time in conducting the review. However, we decided to dedicate this time in order to make our search strategy more inclusive, thus providing a broader overview of the available assessment instruments for dementia in people with ID.

- Lines 167 and 168, not clear why you can not include scales for screening and stratify them for your intended study? (You include the screen in your search strategy?)

We do include screening tools. We are excluding very broad instruments focusing on a general screening for psychiatric symptoms. In the revised paper we include the PAS-ADD as an example. The PAS-ADD is an instrument very well known in our field of research, and will hopefully make this specific exclusion criterium even clearer.

- Include specific measurement properties in the search strategy, such as “validity, reliability, responsiveness…ect.”

In the field of dementia assessment in persons with ID, there are numerous instruments without any proper evaluation of measurement properties. In our first search we want to include/find all instruments, even those without any evaluation (and unfortunately, there are some of those). Therefore, we do not include words referring to psychometric properties in the search strategy as this could lead to an omission/neglect of studies relevant for our review.

- Not clear what the difference is between search one and two, and not clear what happens between the two, will all studies go through the two inclusionary criteria? More clarification is needed. No information on who and how many reviewers there will be per stage, all seems merged.

We divided the description of the two searches by using headings, reading “First search” and “second search”. Everything described under the respective heading refers either to the first or the second search. For example, we listed the inclusion/exclusion criteria for the first search under the heading “first search”, and the inclusion/exclusion criteria for the second search under the heading “second search”.

All papers need to go through both searches. This is also described in the manuscript.

As to the number of reviewers involved, we refer to page line 191-193 of our initial manuscript; which is line 195-197 in the revised manuscript (the version without track changes): “For study selection, one reviewer will exclude duplicates. All remaining records will

be screened and reviewed for eligibility by two team members independently." It can be found at the end of section"search strategy".

- Line 193: There needs to be a method set a-priori for disagreement dissonances, i.e., a third reviewer?

We already specified this method, with the sentence following line 193, reading:" In the case of non-agreement, a third team member will be included in discussion."

- Data extraction: As per the COSMIN guidelines, you need to include the country and the cross-cultural adaptability of the instruments identified in the data extraction.

The COSMIN guidelines include suggestions on which aspects to extract. However, this is dependent on the topic of the study. Most certainly, we will extract "country (language)" in which the instrument was evaluated, as it is also a sample characteristic. In the revised manuscript we added this specifically in the "data extraction" section.

- Line 206: what does important mean? Avoid using terms like this as it is a matter of relevance.

We agree with the reviewer and revised this sentence to be more specific.

- Line 227, "data pooling not possible." Look into this method of qualitative synthesis of measurement properties: https://www.jclinepi.com/article/S0895-4356(06)00174-0/fulltext

We are sorry for the unclarity. We will report a qualitative synthesis of measurement properties. We amended the sentence the reviewer referred to, to read "Quantitative data pooling will probably not be possible."

- Line 246: Zotero can also be used to manage duplication of titles/studies from multiple databases.

We use ZOTERO for managing duplicates and specified this in the revised version.

Regarding the limitations section, study limitations are listed without a discussion on the impact/resolutions of these limitations may have on the outcome of the research.

We think the reviewer is referring to the "study strength and limitation" section. In this section we followed journal style using bullet points to express our study strengths and limitations. This section is not meant to include a discussion. However, we do discuss study limitations, like expected publication bias or quantitative pooling of data, in the main manuscript.

- Overall: A few spelling mistakes, please check the language.

We checked the language and corrected spelling mistakes.

Reviewer: 3
Reviewer Name: Laura Hughes

Institution and Country: University of Sussex, United Kingdom
Competing interests: None declared

Please leave your comments for the authors below
This is a well-written protocol paper for a systematic review of informant based dementia assessment instruments for people with intellectual disabilities.

The content of the review will likely be of value to those researching and working clinically with people with ID and dementia.

Thank you for acknowledging the relevance of our project and for your favorable review.

The introduction clearly lays out the rationale for the review, including discussion for conducting an up-to-date more focussed review on ID and dementia specific instruments. One point to consider is to move or remove the paragraph about COSMIN and CAPs-IDD from the introduction as it already fits well in the methods section and where it is currently situated detracts from the rest of the introduction as it is between the rationale and objectives.

Thank you for this comment. We removed the paragraph. We agree that the respective information is sufficiently described in the methods section.

Methods, analysis and data extraction are well structured and detailed. I am not sure if you need the statement about PPI within the text, the statement about lack of resources does not sit right, it might be more appropriate to state that PPI was not carried out due to the scope of a systematic review of the literature.

We included the PPI statement in within the text in the methods section, as this was demanded by journal guidelines. It is very possible to include patients in the conduction of a systematic review. Unfortunately, we could not involve "patients" (i.e. persons with ID) due to limited resources.

FURTHER CHANGES:

- *Our review was assigned a registration number by PROSPERO (CRD42020181773). We added this information to the manuscript.*
- *We changed one inclusion/exclusion criterium related to language of abstracts. In the previous version of the manuscript, we planned to include studies with an English language abstract, only. We changed this plan, as it is perfectly doable to include all abstracts, irrespective of their language. We can consult team member fluent in other languages, and use translation software.*

VERSION 2 – REVIEW

REVIEWER	Laura Hughes University of Sussex, UK
REVIEW RETURNED	15-Oct-2020

GENERAL COMMENTS	I would like to thank the authors for their revision of the manuscript and addressing reviewer comments.

	The authors have carefully addressed my comments. I have a couple of small amendments that could be addressed prior to publication. In the abstract, could the authors add into the introductory section that they will make recommendations of the most suitable instrument(s). Page 5, line 107: could the authors change this sentence for clarity to state instead "These instrument can be placed into one of three categories" or something similar. Page 6, line 143: remove 'the review protocol'

VERSION 2 – AUTHOR RESPONSE

Reviewer: 3
Comments to the Author
I would like to thank the authors for their revision of the manuscript and addressing reviewer comments.
---Thank you for your time and effort to improve our manuscript.

The authors have carefully addressed my comments. I have a couple of small amendments that could be addressed prior to publication.

In the abstract, could the authors add into the introductory section that they will make recommendations of the most suitable instrument(s).
---We included this information in the introduction of the abstract.

Page 5, line 107: could the authors change this sentence for clarity to state instead "These instrument can be placed into one of three categories" or something similar.
---Thank you for this linguistic suggestion. We changed the respective sentence as suggested.

Page 6, line 143: remove 'the review protocol'
---We changed the respective sentence to be better comprehensible. We did not remove ”the review protocol”, as it is mentioned correctly. However, we subdivided the two parts of the (rather long) sentence, in order to be clearer to the reader. The first sentence is about the review. The second sentence about the review protocol.