

The reliability of recall measurement in assessing migrant reintegration: Evidence from Ethiopia, Somalia, and Sudan

James Dennison^{1,2} 

¹Migration Policy Centre, European University Institute, Fiesole, Italy

²School of Politics, Philosophy, Language and Communication, University of East Anglia, Norwich, UK

Correspondence

James Dennison, Migration Policy Centre, European University Institute, Fiesole, Italy.
Email: james.dennison@eui.eu

Funding information

Leverhulme Trust

Abstract

This article asks whether and under what conditions the use of retrospective or “recall” measurement is reliable. Migration researchers are often forced to retrospectively measure baselines when evaluating the impact of interventions due to the transitory nature of migration, developing country contexts, and hastily assembled policy programmes, a situation exacerbated by Covid-19. This article first theoretically considers the extent to which this approach is reliable and likely to result in biased estimates, as well as its broader advantages, disadvantages, and recommendations for best practice. It then considers the case of the “IMPACT” evaluation of the EU-IOM Joint Initiative for Migrant Protection and Reintegration in which 1774 Ethiopian, Somalia, and Sudanese migrant returnees in 2021 were assessed on a range of reintegration measures, 1095 of whom were measured retrospectively. Regression analyses demonstrate that those measured retrospectively give more negative scores on several “Reintegration Sustainability Scores” in line with some theoretical expectations but contrary to others. However, this—mostly non-statistically significant—effect is largely diminished when the small minority who report finding it difficult to remember the baseline period are removed—suggesting that any retrospective measurement effect results from memory

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Author. *International Migration* published by John Wiley & Sons Ltd on behalf of International Organization for Migration.

bias rather than, for example, consistency bias. No evidence is found to support several theoretically derived interaction effects. Determinants of self-reported memory are demonstrated and recommendations for usage of retrospective measurement are provided, based on these findings.

INTRODUCTION

This article asks to what extent retrospective survey measurement is reliable and to what extent, and under what conditions, it can be optimal. Retrospective survey measurement, also known as “retrospective enumeration”, is characterised by asking individuals about their past—for example, “thinking about your life two years ago, how happy would you say you were on a scale of 0–10?”. These can be used to produce baseline estimates in the calculation of over-time change. This article uses the case of a survey of returnee migrants in Ethiopia, Somalia, and Sudan to investigate retrospective survey measurement. This survey forms one part of the European Union–International Organisation for Migration Joint Initiative for Migrant Protection and Reintegration, launched in December 2016 with €350 million funding from the EU Emergency Trust Fund for Africa (EUTF) to reintegrate migrants in their countries of origin. The programme is expected to contribute to International Organisation for Migration's (IOM) Integrated Approach to Reintegration by addressing returnees' economic, social, and psychosocial needs and to include communities of return in reintegration planning and support.

Scientifically, understanding the extent to which retrospective or recall questions can be used to measure baseline scores is of interest because of ongoing arguments about the extent to which such measures are subject to biases regarding memory, consistency, survival, and social desirability (see review below, including Hipp et al., 2020; Thigpen, 2019). More profoundly, it raises questions about human ability to understand his or her own past and world accurately and objectively or whether such memories and understandings are function to serve cognitive needs in the present. More broadly, comparing retrospective measurement to other forms of baseline measurement—i.e. contemporaneous (also known as prospective) baseline measurement as part of panel data—allows us to consider under what conditions each of the approaches represents the best or “least bad” available approach.

The extent to which retrospective measurement is reliable is also substantively important. Migration, specifically, is likely to remain one of the world's most important and complex political challenges throughout the 21st century, with vast economic and humanitarian consequences and raising profound legal- and rights-based questions for millions of people worldwide. Understanding the actual effects of interventions aimed at improving reintegration outcomes is thus of overwhelming practical importance for advocacy organisations, governments, communicators, policymakers, and those working in politics who want to know what constitutes a sustainable, effective, and value-for-money policy. There are already significant scientific findings on the challenges and determinants of successful reintegration (Arowolo, 2000; Fentaw, 2018; Mercier et al., 2016) and assisted voluntary return and reintegration more generally (Koser & Kuschminder, 2015; Kuschminder, 2022; following Covid-19, Le Coz & Newland, 2021). However, the transitory and international nature of migration, often involving developing countries, means that traditional panel and longitudinal approaches, particularly over lengthy periods, are not always possible. Therefore, identifying if and under what conditions retrospective measurement can be used has plenty of potential uses for policy and programme impact evaluation, regarding the Joint Initiative for Migrant Protections and Reintegration and far beyond, particularly following the disruptive impact of the Covid-19 pandemic. Moreover, testing the reliability of retrospective measurement in the field of migration—as has recently been done in a wide range of other policy areas—may bring to the fore broader and more nuanced lessons for practitioners across disciplines.

The article proceeds as follows. First, the literature on retrospective survey measures is considered, including previous findings on its potential shortcomings, benefits, and best practices. Second, several hypotheses are offered to explain potential discrepancies between retrospective and contemporaneous measurement. Third, the methods and data, including the specific case and survey experiment, are outlined. Fourth, several analyses of the effects of retrospective versus contemporaneous measurement on a range of reintegration indicators, as well as theoretically relevant interaction effects, are presented. Finally, broader ramifications and recommendations based on the analyses are discussed, as well as its shortcomings and next steps for research.

Retrospective survey measurement

Increasingly transnational, fast-moving, and unpredictable crises and transformations—recently including pandemics, migration crises, climate-related disasters, digitalisation and internet threats, terrorism and unconventional warfare, and economic and financial crises—have often left policymakers at various levels of governance scrambling to devise and execute post-facto solutions. Resultant policy programmes—some of which have budgets in the billions of dollars and cross numerous national boundaries—are therefore often hastily assembled. One example of the challenges posed by the measurement and evaluation of these large and reactive programmes is the need to measure change in individuals, ideally requiring *pre-* (also known as baseline or *ex ante*) and *post-* (also known as topline or *ex post*) programme measurement despite no measurement being taken before the programme began. Moreover, even using genuine baseline data to measure change presents numerous problems and biases for researchers, such as panel response, panel fatigue, and panel conditioning and the “Hawthorne effect”, with one result being the need to top-up samples to maintain representativeness (Scott & Alwin, 1998: 16).

Researchers have sought to bypass this issue by retrospectively measuring the baseline; in survey research, notably, by asking individuals to recall their situation *ex-ante* as well as describe their *ex-post* situation at the time of responding to the survey (Hipp et al., 2020). This approach has been recently utilised in assessing the effects of the Covid-19 pandemic and its various public policy responses (Giorgio et al., 2020; Hipp et al., 2020; Li et al., 2020; Ran et al., 2020; SORA, 2020). Researchers investigating long-term socio-economic changes for which there is no available panel data have also turned to retrospective questioning, the data from which is used to produce panel data *post facto* (e.g. Boucher et al., 2007; Fleisher & Wang, 2005; Zhou, 2000). Moreover, researchers investigating developing countries or transitory groups, such as migrants, have needed to make use of retrospective questions, given the difficulty in tracking individuals over time in such contexts (Gibson & Kim, 2010).

Although retrospective measurement has obvious benefits for both researchers and policymakers, it has also been argued to result in notable sources of measurements error, resulting in less accurate responses than contemporaneous measurement (Coughlin, 1990; Hipp et al., 2020; Solga, 2001). First, retrospective questions can only be asked to survivors, both in the literal sense and the broader sense of those that are available for questioning (Scott & Alwin, 1998; although this is also the case for panel data). Second, retrospective questions place high cognitive demands on respondents (Durand et al., 2015; Himmelweit et al., 1978; Yan & Tourangeau, 2007) notably in terms of remembering topics that are or were unimportant to them (Belli, 1998; Bound et al., 2001; Coughlin, 1990; Kennickell & Starr-McCluer, 1997; Peters, 1988; Pina Sánchez et al., 2014; Smith & Thomas, 2003; Teitler et al., 2006) or being unable to accurately date life events, reducing our confidence in temporal ordering (Jacobs, 2002; Paull, 2002) or negatively affecting the accuracy of more volatile events (Gibson & Kim, 2010). Third, individuals have been argued to give biased responses to avoid cognitive dissonance with their analogous contemporaneous self-assessments, narratives, or self-understandings (Barsky, 2002; Jaspers et al., 2009; Schmier & Halpern, 2004; Teitler et al., 2006; Yarrow et al., 1970), or to maintain consistency with contemporaneous social norms and values (a form of social desirability bias; Coughlin, 1990; Himmelweit et al., 1978).

Both of the latter constitute a form of *consistency bias* that would overestimate similarity between the retrospective baseline measurement and the contemporaneous “endline” (or *ex-post* or *post-treatment*)

measurement (Cacioppo, 2002: 130–132; Leising, 2011; Shachar & Eckstein, 2007). Aside from this potential for *underestimating* difference, several biases would also exaggerate the *positivity* with which respondents retrospectively assess their past, notably “rosy retrospection” (Mitchell et al., 1997; Mitchell & Thompson, 1994; Zurbriggen et al., 2021), the related “euphoric recall”, and egocentric bias, whereby, amongst other things, individuals overestimate their own outcomes in hindsight (see also “hindsight bias”; Connolly & Bukszar, 1990; Roese & Vohs, 2012). Alternatively, however, there may be a *negative* effect of retrospective measurement if the treatment leads individuals to reappraise *ex post* their subjective assessment of their situation *ex ante*, i.e. their standards are changed by the treatment so that they judge their previous situation more negatively than they would have at the time, a form of systematic measurement error whereby the treatment affects the baseline (Cochran, 1968; Saris & Revilla, 2015).

However, empirical evidence remains too mixed to entirely rule out retrospective measurement, certainly in comparison to other approaches. Thigpen (2019: 453) finds that ‘the passage of time has a weak, uncertain influence on recollection’ and concludes that ‘the recalled answers display a sufficient degree of correspondence to prospectively collected (i.e. contemporaneously collected) response to have faith in analyses using retrospective survey data regarding traits within the past five years or less.’ Similarly, Peters (1988) finds substantive agreement between retrospective and panel measurement approaches, with errors found to increase with more cognitively demanding questions, and Pierret (2001: 439) finds that “coefficients ... did not change greatly” between the two approaches. Smith and Thomas (2003: 47) compare panel and retrospective approaches for measuring migration histories, concluding that retrospective questions ‘can elicit useful information if we can isolate those events which people remember from those which they forget’, which they suggest can be more easily done by using retrospective on more recent events (with the last 2 year years) of longer duration (over 6 months), linking questions to other high salience life events (e.g. marriage, having children), asking multiple members of the same household the same retrospective questions, and having interviewers record the quality of the interview. Overall, they suggest that ‘the current presumption against the use of long-term recall questions in field surveys ignores a potentially rich source of data’. Muggenburg (2021) also makes several suggestions on the content of questions and survey design that are likely to reduce bias in retrospective measurement, with theoretical justifications for each. None of these studies validate data from either method so that the accuracy of retrospective data remains a source of ‘continued debate’ (Gibson & Kim, 2010: 687). Overall, Scott and Alwin (1998: 16, 23; see also Muggenburg 2021, for similar conclusion) note that both retrospective and prospective approaches contain several potential sources of bias so that ‘if both prospective and retrospective data have merits, then the ideal may point to a middle ground solution’ so that ‘retrospective measurement is, for most social researchers, an indispensable tool’.

Hypotheses

Hypothesis 1. Retrospective measurement of reintegration has a *positive* effect on baseline reintegration metrics compared to contemporaneous measurement (*consistency* and *rosy retrospective* biases).

Hypothesis 2. Retrospective measurement of reintegration has a *negative* effect on baseline reintegration metrics compared to contemporaneous measurement (*systematic measurement error*).

Hypothesis 3. Self-assessed memory of the time that the baseline refers negatively affects the effect of retrospective measurement on reintegration metrics (*memory bias*).

Moreover, several interaction effects are theoretically plausible based on the literature whereby the difference in migrant reintegration scores between those asked retrospectively and those asked contemporaneously increases, namely:

Hypothesis 4a. The *duration* of time between the time of questioning and the time of return.

Hypothesis 4b. A *lower education* level of the respondent.

The IMPACT study, methods, and data

To test the above hypotheses, this article uses data from the “IMPACT” study, which evaluated the EU-IOM Joint Initiative for Migrant Protection and Reintegration (henceforth, the EU-IOM Joint Initiative). The EU-IOM Joint Initiative comes in the context of a change in thinking around migrant re-integration. Whereas re-integration was formerly conceived of—or at least acted upon—in terms of cash for returnees, larger and more developed programmes since the mid-2010s increasingly conceive of reintegration programmes in structural terms involving governments and communities, in addition to individuals. Similarly, the stated objectives of such programmes, and thus the ways in which their impacts are measured, have turned away from purely economic indicators to also include social and psychosocial measures. Overall, practically, this has meant that these more sophisticated and ambitious reintegration programmes have left their administrators, such as the IOM, responsible for delivering a broad range of services to returning migrants to assist their reintegration across numerous aspects of life, including business start-up support, training, education, and job placement, financial support and management, improved access to services, and psychosocial support including community integration, support groups, and mentorship.

The EU-IOM Joint Initiative was launched in December 2016 with €350 million in funding from the EU Emergency Trust Fund for Africa (EUTF) and claims to be ‘the first comprehensive programme to save lives, protect and assist migrants along key migration routes in Africa’.¹ It does so by assisting migrants who decide to return to their countries of origin through cooperation between partnering state and non-state actors that aims to support both migrants and their communities across 26 African countries. The stated objectives of the Joint Initiative include providing direct assistance and enabling the assisted voluntary return of migrants stranded along the migration routes and supporting the reintegration process of returning migrants in an integrated approach which addresses economic, social, and psychosocial dimensions and fosters the inclusion of communities of return, amongst other objectives related to outbound migrants.

Overall, the programme is considered by the IOM to be a flagship and first attempt to apply an “integrated approach to reintegration” in contrast with ‘traditional’ return and reintegration programmes (for detailed information on what the programme does and the IOM’s reintegration approach, see IOM, 2019). However, the programme was also implemented in a rush under strong political pressure (Dennison & Geddes, 2019). To measure the EU-IOM Joint Initiative, and reintegration programmes more generally, the IOM’s 2017 “Measure” project defined a framework for the measurement of reintegration outcomes of returning migrants: Reintegration Sustainability Score (RSS) and a methodology to compute the RSS with the data gathered through the survey questionnaire (Samuel Hall and IOM, 2017). Guidelines on the usage of this measurement framework recommend the administration of the RSS twice: the first time to establish a ‘baseline’ measurement—around 3 months after return—and a second time to establish an ‘endline’ measurement—at least 1 year after return.

Building on this work, the IMPACT study, starting in 2019, sought to assess the impact of the reintegration assistance provided by the EU-IOM Joint Initiative, although restricted to Ethiopia, Somalia, and Sudan. It is the

data from that study that this article relies upon (see methodological report, Eager et al., 2020). The IMPACT study used a quasi-experimental design whereby reintegration outcomes of members of a treatment group composed of programme beneficiaries are systematically compared with the (re)integration outcomes of members of a “calibration” (i.e. control) group. This latter group is composed of non-migrant individuals who were not assisted by the programme and are matched to one member of the treatment group based on a defined set of socio-demographical criteria including residing in the same communities where the beneficiaries live. The RSS in this study was modified through a comprehensive review process, which resulted in an enhanced tool called RSS+. The RSS+ includes a set of ‘retrospective’ questions (not included in RSS) which aim at allowing the computation of a ‘retrospective-baseline’ RSS score in case baseline data is not available for the respondent interviewed.

The RSS+ study is an evolution of the original RSS questionnaire and, crucially for this article, contains retrospective questions as well as additional information on the respondent’s understanding of retrospective questions. The introduction of retrospective-baseline questions into RSS+ was motivated by both practical and methodological considerations. On the practical side, the outset of the pandemic caused a significant decrease in the influx of new returns, making it therefore impossible to achieve the targeted sample size with new returns alone, given that baseline data is available only for a small fraction of the beneficiaries returned before the launch of the IMPACT study. On the methodological side, provided proper design and methodological precautions are in place, it might be argued that retrospective enumeration may present possible benefits in terms of bias reduction and elaboration of the questions by the respondents.

Data

Data continues to be collected for the IMPACT study at the time of writing, including a full control group that did not receive IOM reintegration assistance. Such data will continue to be collected until around the end of 2023. However, the 1774 observations collected by the IMPACT Study by early 2021 can already be used to test the above hypotheses. Although this survey included many dozen variables, for the purposes of this article, only ten independent variables and four dependent variables are utilised (the comprehensive Reintegration Sustainability Score and its three constituent parts: economic, psychosocial, and social). The variable names, their possible responses, and the distribution of those responses for the purposes of this study are shown in Table 1. It is worth noting that the large, complex nature of RSS data and the migration context often results in missing variables, attrition, etc., although in this case the regression models of RSS scores contain 1648 observations, a relatively minor drop from the full sample. Most of these variables and the justification for their inclusion are relatively self-explanatory. *Psychological problems* records responses to the question “Do you often suffer from any of the following? Feeling angry; Feeling sad; Feeling afraid; Feeling stressed; Feeling lonely; Feeling low self-worth; Difficulty concentrating” and is included as a control variable in case such symptoms lead to memory problems. *Memory* records responses to the question “When asked about things related to the past (3 months after return), how easy or difficult was it to remember your situation back then and answer the questions?” and is included to test whether such responses affect the effect of the form of baseline measurement (i.e. whether they impair the accuracy of retrospective measurement). Because this is only asked to those individuals responding retrospectively, it is not included in the main regression analyses.

Methods

Hypothesis 1 and 2 are tested by comparing baseline scores between those who were and were not retrospectively measured in two ways. Most naively, this can be done by simply comparing the four mean baseline RSS scores amongst those measured contemporaneously and those measured retrospectively. However, there

TABLE 1 Descriptive variables.

Discrete variables			
Variable	Response	N	%
Origin (return) country	Ethiopia	601	33.9
	Somalia	206	11.6
	Sudan	967	54.5
Sex	Female	292	16.5
	Male	1481	83.5
Baseline	Contemporaneous	679	38.3
	Retrospective	1095	61.7
Education	None	156	8.3
	Primary	795	41.9
	Secondary / religious / vocational	736	38.8
	Tertiary	209	11.0
Psychological problems	Never	771	44.1
	Rarely	590	33.7
	Sometimes	197	11.3
	Often	169	9.7
	Very often	23	1.3
Interview type	In person	468	26.4
	Phone-based	1308	73.7
Primary applicant	Non-primary respondent (e.g. family member responding)	103	5.8
	Primary respondent	1676	94.2
Memory	Difficult to remember 3 months after return	73	6.7
	Neither easy nor difficult	154	14.1
	Easy to remember	865	79.2
Variable	Mean	Min	Max
Reintegration sustainability scores			
RSS Comprehensive	0.60	0.23	0.92
RSS Economic	0.43	0.09	0.93
RSS Psycho-social	0.77	0.13	1
RSS Social	0.56	0.08	1
Continuous variables			
Age at arrival	29	14	87
Days since baseline	371 (inc. 0s); 601 (not incl. 0s)	0 for contemp. baseline; 299 for retro. baseline	1614

may be several confounding variables, particularly since the mode of baseline measurement was not randomly distributed for practical reasons outlined above, including the rapid rollout of the IMPACT study and the disruptive effects of the Covid-19 pandemic. As such, linear regression models can be analysed whereby one's

continuous RSS score is the dependent variable, their form of baseline measurement is the key independent variable of interest, and several other control variables are included regarding one's socio-demographics (gender, education, psychological health, age, and their national context; Arowolo, 2000; Fentaw, 2018; Koser & Kuschminder, 2015; Kuschminder, 2022; Le Coz & Newland, 2021; Mercier et al., 2016), and measurement issues such as time between the baseline and its measurement, whether the measurement was done face-to-face or via phone (Holbrook et al., 2003), and whether it is done with the migrant or member of their family or cohabitee (Davin et al., 2019). Propensity Score Matching (PSM) is not used given the strong recent criticism of this approach when—as in this case—one is testing the effects of a non-randomly distributed treatment; namely that it 'accomplishes the opposite of its intended goal – thus increasing imbalance, inefficiency, model dependence, and bias' (King & Nielsen, 2019: 435) and 'there is no way to know in advance whether the method will work' (Peikes et al., 2008: 62; for further discussion on PSM criticism, debates, and under what conditions it may be used see Guo et al., 2020; Smith & Todd, 2001; Wang, 2021).

Importantly, however, given that several methodological sources of bias stemming from retrospective measurement are related to memory, it is also important to see if our results vary by one's self reported ease of memory of the time period in which the baseline was taken (Hypothesis 3). This is done simply by repeating the four regression analyses (one for each of our RSS measures) with those not reporting that remembering that time was easy (21 per cent of retrospective responders) being excluded from the analysis. Finally, Hypothesis 4a and 4b can be tested by running the same regression analyses, but including interaction effects for level of education and for the duration of time between the baseline measurement and the time period to which it refers (i.e. 3 months after return).

Analyses

The simple mean RSS scores of those whose baseline was measured contemporaneously and those whose baseline was measured retrospectively are shown in Table 2. We can clearly see only minor differences between the two groups across all four of the RSS scores. Overall, these initial averages provide most support for the null hypothesis.

This is further investigated by considering variation in the above responses according to days since measurement, with the latter divided into five groups for ease of comparison in Table 3 below. Overall, we see no evidence of systematic variation, either positively or negatively, according to days since measurement.

In Table 4, our regression analyses are presented. Model 1 shows the predictors of the comprehensive RSS score, on which retrospective measurement has a non-statistically significant negative effect of 0.032 (the full scale is between 0 and 1), an effect which is larger than, for example, education. This finding offers some support for Hypothesis 2, although without statistical significance we cannot be certain about the finding in the population. However, notably, as shown in model 2, the effect of retrospective measurement falls greatly when the 20 per cent of those measured retrospectively who reported finding it not easy to remember the time-period are removed from the analysis. This suggests that the negative effect observed in model 1 is partially a result of

TABLE 2 Mean RSS scores by form of baseline measurement.

Variable	Contemporaneous mean	Retrospective mean
RSS Comprehensive	0.61	0.59
RSS Economic	0.43	0.43
RSS Psycho-social	0.78	0.77
RSS Social	0.56	0.55

TABLE 3 Mean RSS scores by form of baseline measurement and days since retrospective measurement.

Variable	Contemp. Baseline mean	Retrospective baseline mean			
		1st (299–460 days)	2nd (461–507 days)	3rd (508–792 days)	4th (793–1615 days)
RSS Comprehensive	0.61	0.61	0.58	0.58	0.61
RSS Economic	0.43	0.46	0.40	0.40	0.44
RSS Psycho-social	0.78	0.79	0.74	0.74	0.80
RSS Social	0.56	0.53	0.56	0.59	0.54

memory bias, supporting Hypothesis 3. The remaining models 3–8 have mixed results, generally showing that the effect of retrospective measurement is not statistically significant—with the effects largest in the case of psychological indicators (perhaps reflecting their more subjective basis) and weakest in the case of social and especially economic indicators. Two coefficients are statistically significant, negatively on the psychosocial RSS (model 6, supporting Hypothesis 2) and positively on the social RSS (model 8, supporting Hypothesis 1). However, importantly and shown in models 4, 6, and 8, this effect becomes considerably smaller when those who report not finding it easy to remember are removed and, in the case of the effects of the economic and social RSSs, becomes non-statistically significant in all cases. Some of the control variables also display consistent and theoretically interesting effect. When the primary respondent to the survey is the returning migrant her- or himself, the RSS score increases in every model, except the two relating to social services (perhaps reflecting the community nature of these indicators). The use of phone interview rather than in person increases the RSS score in every case, while there is considerable evidence that a higher age has the same effect. Overall, we find the strongest evidence to support the null hypothesis, which in the case of the comprehensive RSS score becomes stronger when those who cannot remember are removed.

Moving on to our test of Hypothesis 4a or 4b displayed in Table 5, we can see that neither days since the baseline (models 1 and 2) nor education level (models 3 and 4) have either large or statistically significant effects on the effect of retrospective measurement on the comprehensive RSS. As such, we find little evidence to support Hypothesis 4a or 4b so, theoretically, memory bias does not seem to be a function of lack of education or time since the event. Models 5 and 6 also show how the effect of retrospective measurement varies by country. We can see that the negative effect of retrospective measurement is exacerbated by being in Somalia rather than Ethiopia and reduced by being in Sudan rather than Ethiopia. Notably, it is only when this interaction is included do we see a statistically significant main effect of retrospective measurement on the comprehensive RSS.

Determinants of retrospective memory

Given the clear importance of self-reported retrospective memory in determining the effect of retrospective measurement on one's RSS score, it is worth considering further the determinants of self-reported retrospective memory. To do so, a dichotomous variable of retrospective memory is produced (0 for those saying it is easy to remember 3 months after return and 1 for those either stating that it is difficult to do so or neither difficult nor easy). Along with the same socio-demographic and measurement control variables as the earlier regressions, the resultant regression analysis predicting ease of memory is shown in Table 6. Notably, none of gender, education, age, or days since the baseline measurement influence one's memory. Instead, we see that being in Sudan, rather than Ethiopia, decreases the chance that one says it is not easy to remember whereas, unsurprisingly, having psychological problems increases the chance. Similarly, doing the interview over the

TABLE 4 Regression analyses of determinants of RSS scores, by inclusion and exclusion of those who self-report as not finding it easy to remember the baseline time period.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RSS comp	RSS comp	RSS comp	RSS econ	RSS econ	RSS psych	RSS psych	RSS social	RSS social
		Easy to remember retro only		Easy to remember retro only		Easy to remember retro only		Easy to remember retro only
Retrospective	-0.032 (0.014)	-0.007 (0.003)	0.002 (0.012)	0.025 (0.013)	-0.061 (0.026)	-0.030* (0.009)	-0.005 (0.007)	0.014*** (0.001)
Male	-0.000 (0.008)	-0.002 (0.008)	0.013* (0.004)	0.009 (0.006)	0.008 (0.014)	0.007 (0.015)	-0.031*** (0.002)	-0.032*** (0.002)
Country (ref: Ethiopia) Somalia	0.053*** (0.004)	0.012 (0.005)	-0.089*** (0.009)	-0.119*** (0.009)	-0.035** (0.004)	-0.084** (0.009)	0.156*** (0.007)	0.135*** (0.005)
Sudan	-0.016 (0.009)	-0.004 (0.003)	-0.101** (0.013)	-0.097*** (0.010)	0.062* (0.017)	0.076** (0.009)	0.064* (0.017)	0.070* (0.016)
Psych problems	0.049** (0.008)	0.029*** (0.002)	0.020 (0.012)	0.004 (0.008)	0.134** (0.016)	0.111*** (0.010)	0.004 (0.017)	-0.004 (0.019)
Education (ref: none)	-0.041** (0.005)	-0.042*** (0.002)	-0.034*** (0.003)	-0.035*** (0.001)	-0.078** (0.011)	-0.079** (0.009)	-0.007 (0.003)	-0.007 (0.005)
Primary	0.002 (0.012)	0.000 (0.012)	-0.012 (0.004)	-0.016* (0.004)	-0.033 (0.013)	-0.034 (0.015)	0.053 (0.034)	0.052 (0.034)
Secondary	0.009 (0.006)	0.006 (0.005)	-0.008 (0.003)	-0.015 (0.009)	-0.030** (0.005)	-0.028* (0.007)	0.077** (0.017)	0.074** (0.015)
Tertiary	0.032** (0.006)	0.030** (0.006)	0.000 (0.011)	-0.005 (0.012)	-0.021 (0.008)	-0.022 (0.009)	0.135** (0.021)	0.134** (0.019)
Age at arrival	0.001 (0.000)	0.001** (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.001 (0.000)	0.001** (0.000)	0.001** (0.000)

TABLE 4 (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	RSS comp	RSS comp	RSS econ	RSS econ	RSS psych	RSS psych	RSS social	RSS social
		Easy to remember retro only						
Days since baseline	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
Phone interview	0.031* (0.008)	0.049** (0.006)	0.033* (0.010)	0.047** (0.008)	0.033 (0.014)	0.055** (0.012)	0.032** (0.007)	0.043** (0.006)
Primary applicant	0.024* (0.008)	0.024* (0.007)	0.025 (0.011)	0.031** (0.007)	0.041* (0.014)	0.044** (0.007)	0.000 (0.004)	-0.008 (0.013)
Constant	0.564*** (0.037)	0.540*** (0.033)	0.413*** (0.033)	0.396*** (0.029)	0.773*** (0.048)	0.738*** (0.039)	0.438*** (0.028)	0.424*** (0.035)
Observations	1648	1456	1648	1456	1648	1456	1648	1456
R-squared	0.331	0.344	0.168	0.160	0.524	0.568	0.138	0.143

Note: Standard errors in parentheses and clustered by country.

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

TABLE 5 Regression models testing interaction effects between retrospective measurement and days since baseline, education, and country.

	(1)	(2)	(3)	(4)	(5)	(6)
	RSS comp	RSS comp Easy to remember retro only	RSS comp	RSS comp Easy to remember retro only	RSS comp	RSS comp Easy to remember retro only
Retrospective	-0.032 (0.014)	-0.007 (0.003)	-0.031 (0.025)	-0.003 (0.010)	-0.058** (0.008)	-0.030* (0.009)
Retrospective x days	-0.000 (0.000)	-0.000 (0.000)				
Retrospective x primary			0.004 (0.014)	-0.001 (0.010)		
Retrospective x secondary			-0.003 (0.021)	-0.004 (0.009)		
Retrospective x tertiary			-0.006 (0.017)	-0.017 (0.011)		
Retrospective x Somalia					-0.052** (0.005)	-0.027* (0.007)
Retrospective x Sudan					0.042** (0.007)	0.031** (0.005)
Male	0.053*** (0.004)	0.012 (0.005)	0.052*** (0.003)	0.011 (0.005)	0.074*** (0.005)	0.030** (0.005)
Country (ref: Ethiopia)						
Somalia	-0.016 (0.009)	-0.004 (0.003)	-0.017 (0.007)	-0.005 (0.002)	0.027** (0.003)	0.017** (0.003)
Sudan	0.049** (0.008)	0.029*** (0.002)	0.048** (0.008)	0.028*** (0.002)	0.032** (0.003)	0.020** (0.004)

TABLE 5 (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)
	RSS comp	RSS comp	RSS comp	RSS comp	RSS comp	RSS comp
		Easy to remember retro only				
Psych problems	-0.041** (0.005)	-0.042*** (0.002)	-0.041** (0.004)	-0.042*** (0.002)	-0.040** (0.005)	-0.041*** (0.003)
Education (ref: none)						
Primary	0.002 (0.012)	0.000 (0.012)	-0.001 (0.019)	0.001 (0.017)	0.003 (0.013)	0.001 (0.014)
Secondary	0.009 (0.006)	0.006 (0.005)	0.011 (0.012)	0.008 (0.009)	0.006 (0.007)	0.004 (0.006)
Tertiary	0.032** (0.006)	0.030** (0.006)	0.035 (0.015)	0.038* (0.012)	0.032** (0.007)	0.029* (0.007)
Age at arrival	0.001 (0.000)	0.001** (0.000)	0.001 (0.000)	0.001** (0.000)	0.001* (0.000)	0.001** (0.000)
Days since baseline	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000* (0.000)	0.000 (0.000)
Phone interview	0.031* (0.008)	0.049** (0.006)	0.032* (0.008)	0.050** (0.006)	0.022** (0.002)	0.041*** (0.003)
Primary applicant	0.024* (0.008)	0.024* (0.007)	0.024* (0.008)	0.024* (0.007)	0.021 (0.011)	0.022 (0.009)
Constant	0.564*** (0.037)	0.540*** (0.033)	0.563*** (0.040)	0.536*** (0.036)	0.587*** (0.027)	0.559*** (0.029)
Observations	1648	1456	1648	1456	1648	1456
R-squared	0.331	0.344	0.332	0.345	0.350	0.352

Note: Standard errors in parentheses; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

TABLE 6 Determinants of not finding it easy to remember the time of 3 months after return.

	(1)
	Not easy to remember
Male	0.001 (0.121)
Country (ref: Ethiopia)	
Somalia	0.084 (0.226)
Sudan	-1.782*** (0.389)
Psych problems	0.163* (0.191)
Education (ref: none)	
Primary	-0.145 (0.086)
Secondary	-0.128 (0.431)
Tertiary	-0.277 (0.433)
Age at arrival	0.016 (0.012)
Days since baseline	-0.000 (0.001)
Phone interview	0.835*** (0.039)
Primary applicant	-0.102 (0.460)
Constant	-2.265*** (0.939)
Observations	988

Note: Standard errors in parentheses and clustered by country.

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

phone, rather than face-to-face also increases one's chance of saying that they cannot remember the period easily.

Discussion and recommendations

This article considered to what extent retrospective survey measurement is reliable and what are its advantages, disadvantages, and best practices. The transitory nature of migration and the developing country contexts for—at times, hastily assembled—policy programmes often make panel and randomised control trial designs impractical when evaluating migration reintegration projects, a situation recently exacerbated by Covid-19. Migration researchers therefore often turn to retrospective baseline measures, an approach that theoretically may result in measurement biases in

both positive and negative directions. As such, data from the IMPACT evaluation study of the EU-IOM Joint Initiative for Migrant Protections and Reintegration was tested, in which 1774 Ethiopian, Somalia, and Sudanese migrant returnees in 2021 were assessed on a range of reintegration measures, 1095 of whom were measured retrospectively. Regression analyses demonstrate that those measured retrospectively give more negative scores on several “Reintegration Sustainability Scores”—though these effects were mostly not statistically significant—in line with theoretical expectations regarding measurement error, but contrary to several theorised biases. However, this effect weakened considerably when the small minority who report finding it difficult to remember the baseline period are removed—suggesting that any retrospective measurement effect partially results from memory bias rather than, for example, consistency bias. No evidence was found to support several theoretically derived interaction effects.

This article therefore makes several contributions. First, it is demonstrated that the effect of giving retrospective than contemporaneous baseline measurements is negative where it exists, contrary to our expectations from at least three well-documented cognitive biases but consistent with the possibility of endogeneity. Second, it is shown that this effect partially disappears when the small minority of retrospective respondents who state that it was not “easy to remember” the time of the baseline are removed. As such it is argued that much of the retrospective bias comes from memory and so, third, it is recommended that researchers thus ensure that a measure is included on self-reported memory when using retrospective measurement, notwithstanding the risks of non-randomly distributed memory of the baseline. Fourth, several commonly theorised mediating effects of the effect of retrospective measurement—such as education level, time elapsed since the baseline—are tested without supporting evidence. Fifth, the determinants of retrospective memory are also tested, showing that socio-demographics have little effect whereas the mode of interview, psychological problems, and country of origin do. Overall, retrospective measurement has significant advantages over panel approaches—notably in terms of efficiency—since panel approaches are subject to attrition and other biases that are especially problematic in dealing with transient populations. This article thus validates recent methodological recommendations to utilise both forms of baseline measurement in impact evaluations.

As such, because retrospective measurement has significant practical and logistical advantages over panel approaches, notably in terms of efficiency, subject as panel approaches are to attrition and other biases that are especially problematic in dealing with transient populations, this article recommends that researchers ideally utilise both forms of baseline measurement in impact evaluations. When using retrospective measurement, however, it is vital to:

1. Simultaneously gather data on self-reported ease-of-memory of the time being measured.
2. Systematically test for variation in the results according to self-reported memory.
3. Because phone interviews are found to increase self-reported problems in memory, it is recommended that face-to-face interviews are especially prioritised.

This article is necessarily limited and represents a first attempt, both theoretically and empirically, to consider the appropriate use of retrospective measurement in migration research. The lack of randomisation—necessary in the case of this programme—between those measured retrospectively and those measured contemporaneously may have biased the results if there were differences between the two groups that co-vary with any factors affecting the reintegration measures—though the, primarily, null findings reduce the likelihood that this was the case. Furthermore, further research should be made into how retrospective measurement accuracy diminishes over time, regardless of self-assessed memory of the period. This study found that people's reported memory is not affected by the time since the period being remembered, contrary to expectations regarding memory bias (although in line with Smith and Thomas's (2003) expectation of accuracy within 2 years since the mean for this study was 602 days with a median of 509). This by no means suggests that memory bias will not increase at some point or in some circumstances though, with the lack of interaction effect perhaps reflecting the relatively short time period, the salience of the time period of the baseline measurement in an individual's life, or biases increasing

but “cancelling out” over time (some combination of survivor bias, cognitive demand, cognitive consonance, rosy retrospection, euphoric recall, and egocentric bias, with some of these having theoretically plausible contrary effects).

Future research should utilise the IMPACT study's growing body of data, including the future inclusion of a control group, to produce genuine randomised control trials from which the effect of retrospective measurement can be further tested. Similarly, as the number of individuals increases, it will be possible to effectively test how participation in various reintegration activities (such as starting a microbusiness, done by 410 participants in this initial sample) affected—above all—the RSS scores, but also secondarily the effect of the type of baseline measurement. Furthermore, given the seeming centrality of memory bias to problems of retrospective measurement—and the inability of this article to find socio-demographic determinants of it—it is vital to further investigate the causes of variation in this memory variable.

PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/imig.13177>.

DATA AVAILABILITY STATEMENT

The full dataset used for this article contains confidential information. However, a reduced version can be provided upon request for replication purposes.

ORCID

James Dennison  <https://orcid.org/0000-0003-3090-7124>

ENDNOTE

¹ <https://www.migrationjointinitiative.org/about-eu-iom-joint-initiative> accessed on 10 February 2021.

REFERENCES

- Arowolo, O.O. (2000) Return migration and the problem of reintegration. *International Migration*, 38(5), 59–82.
- Barsky, A.J. (2002) Forgetting, fabricating, and telescoping. *Archives of Internal Medicine*, 162(9), 981–984.
- Belli, R.F. (1998) The structure of autobiographical memory and the event history calendar: potential improvements in the quality of retrospective reports in surveys. *Memory*, 6(4), 383–406.
- Boucher, S., Smith, A., Taylor, J.E. & Yúnez-Naude, A. (2007) Impacts of policy reforms on the supply of Mexican labor to U.S. farms: new evidence from Mexico. *Review of Agricultural Economics*, 29, 4–16.
- Bound, J., Brown, C. & Mathiowetz, N. (2001) Measurement error in survey data. In: *Handbook of econometrics*. Amsterdam: Elsevier, pp. 3705–3843.
- Cacioppo, J. (2002) *Foundations in social neuroscience*. Cambridge, Mass: MIT Press, pp. 130–132.
- Cochran, W.G. (1968) Errors of measurement in statistics. *Technometrics*, 10(4), 637–666.
- Connolly, T. & Bukszar, E.W. (1990) Hindsight bias: self-flattery or cognitive error? *Journal of Behavioral Decision Making*, 3, 205–211.
- Coughlin, S.S. (1990) Recall bias in epidemiologic studies. *Journal of Clinical Epidemiology*, 43(1), 87–91.
- Davin, B., Joutard, X. & Paraponaris, A. (2019) "If you were me": Proxy respondents' biases in population health surveys.
- Dennison, J. & Geddes, A. (2019) A rising tide? The salience of immigration and the rise of anti-immigration political parties in Western Europe. *Political Quarterly*, 90(1), 107–116.
- Durand, C., Deslauriers, M. & Valois, I. (2015) Should recall of previous votes be used to adjust estimates of voting intention? Survey insights: methods from the field, weighting: practical issues and 'how to' approach. Available from: surveyinsights.org/?p=3543
- Eager, R., Pinney, A., Loevinsohn, M., Sandri, E., EvansGutierrez, L., Kindler, B. et al. (2020) Methodological Report: IMPACT - Impact Evaluation of the EU-IOM Joint Initiative for Migrant Protection and Reintegration in the Horn of Africa region. Available from: https://eastandhornofafrica.iom.int/sites/g/files/tmzbd1701/files/documents/iom_methodological_report_final_20102020.pdf [Accessed 15 February 2022]

- Fentaw, A.E. (2018) Ethiopian returnee women from Arab countries: challenges of successful reintegration. *African and Black Diaspora: An International Journal*, 11(1), 33–50.
- Fleisher, B. & Wang, X. (2005) Returns to schooling in China under planning and reform. *Journal of Comparative Economics*, 33, 265–277.
- Gibson, J. & Kim, B. (2010) Non-classical measurement error in long-term retrospective recall surveys. *Oxford Bulletin of Economics and Statistics*, 72, 687–695.
- Giorgio, E.D., Riso, D.D., Mioni, G. & Cellini, N. (2020) The interplay between mothers' and children behavioural and psychological factors during COVID-19: an Italian study. *European Child & Adolescent Psychiatry*, 30, 1401–1412. Available from: <https://doi.org/10.31234/osf.io/dqk7h>
- Guo, S., Fraser, M. & Chen, Q. (2020) Propensity score analysis: recent debate and discussion. *Journal of the Society for Social Work and Research*, 11(3), 463–482.
- Himmelweit, H.T., Biberian, M.J. & Stockdale, J. (1978) Memory for past vote: implications of a study of bias in recall. *British Journal of Political Science*, 8(3), 365–375.
- Hipp, L., Bünning, M., Munnes, S. & Sauer mann, A. (2020) Problems and pitfalls of retrospective survey questions in COVID-19 studies. In: *Survey research methods*, Vol. 14. Konstanz: European Survey Research Association, pp. 109–114.
- Holbrook, A.L., Green, M.C. & Krosnick, J.A. (2003) Telephone versus face-to-face interviewing of national probability samples with long questionnaires: comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, 67(1), 79–125.
- IOM. (2019) Reintegration Handbook - Practical guidance on the design, implementation and monitoring of reintegration assistance. Available from: <https://publications.iom.int/books/reintegration-handbook-practical-guidance-design-implementation-and-monitoring-reintegration> [Accessed 15 February 2022]
- Jacobs, S. (2002) Reliability and recall of unemployment events using retrospective data. *Work, Employment and Society*, 16, 537–548.
- Jaspers, E., Lubbers, M. & Graaf, N.D.D. (2009) Measuring once twice: an evaluation of recalling attitudes in survey research. *European Sociological Review*, 25(3), 287–301.
- Kennickell, A. & Starr-McCluer, M. (1997) Retrospective reporting of household wealth: evidence from the 1983–89 survey of consumer finances. *Journal of Business and Economic Statistics*, 15, 452–463.
- King, G. & Nielsen, R. (2019) Why propensity scores should not be used for matching. *Political Analysis*, 27(4), 435–454.
- Koser, K. & Kuschminder, K. (2015) *Comparative research on the assisted voluntary return and reintegration of migrants*, Vol. 343. Geneva: International Organization for Migration.
- Kuschminder, K. (2022) Reintegration strategies. In: *Handbook of return migration*. Cheltenham: Edward Elgar Publishing, pp. 200–211.
- Le Coz, C. & Newland, K. (2021) *Rewiring migrant returns and reintegration after the COVID-19 shock*. Washington DC: Migration Policy Institute.
- Leising, D. (2011) The consistency bias in judgments of one's own interpersonal behavior: two possible sources. *Journal of Individual Differences*, 32(3), 137–143.
- Li, Y.K., Peng, S., Li, L.-Q., Wang, Q., Ping, W., Zhang, N. et al. (2020) Clinical and transmission characteristics of COVID-19 - a retrospective study of 25 cases from a single thoracic surgery department. *Current Medical Science*, 40(2), 295–300. Available from: <https://doi.org/10.1007/s11596-020-2176-2>
- Mercier, M., David, A., Mahia, R. & De Arce, R. (2016) Reintegration upon return: insights from Ecuadorian returnees from Spain. *International Migration*, 54(6), 56–73.
- Mitchell, T. & Thompson, L. (1994) A theory of temporal adjustments of the evaluation of events: rosy prospectation & rosy retrospection. In: Stubbart, C., Porac, J. & Meindl, J. (Eds.) *Advances in managerial cognition and organizational information-processing*, Vol. 5. Greenwich, CT: JAI press, pp. 85–114.
- Mitchell, T., Thompson, L., Peterson, E. & Cronk, R. (1997) Temporal adjustments in the evaluation of events: the "rosy view". *Journal of Experimental Social Psychology*, 33(4), 421–448.
- Müggenburg, H. (2021) Beyond the limits of memory? The reliability of retrospective data in travel research. *Transportation Research Part A: Policy and Practice*, 145, 302–318.
- Paull, G. (2002) Biases in the reporting of labour market dynamics. Institute for Fiscal Studies Working Paper No. 02/10.
- Peikes, D.N., Moreno, L. & Orzol, S.M. (2008) Propensity score matching: a note of caution for evaluators of social programs. *The American Statistician*, 62(3), 222–231.
- Peters, E. (1988) Retrospective versus panel data in analyzing lifecycle events retrieved from. *Journal of Human Resources*, 23(4), 488–513. Available from: <http://www.jstor.org/stable/10.2307/145810>
- Pierret, C. (2001) Event history data survey recall: an analysis of the National Longitudinal Survey of Youth 1979 recall experiment. *Journal of Human Resources*, 36, 439–466.

- Pina Sánchez, J., Koskinen, J. & Plewis, I. (2014) Measurement error in retrospective work histories. *Survey Research Methods*, 8(1), 43–55.
- Ran, L., Chen, X., Wang, Y., Wu, W., Zhang, L. & Tan, X. (2020) Risk factors of healthcare workers with corona virus disease 2019: a retrospective cohort study in a designated hospital of Wuhan in China. *Clinical Infectious Diseases*, 71, 2218–2221. Available from: <https://doi.org/10.1093/cid/ciaa287>
- Roose, N.J. & Vohs, K.D. (2012) Hindsight bias. *Perspectives on Psychological Science*, 7(5), 411–426.
- Samuel Hall think tank and IOM. (2017) Setting standards for an integrated approach to reintegration, commissioned by IOM and funded by DFID. Available from: https://www.iom.int/sites/g/files/tmzbd1486/files/our_work/DMM/AVRR/IOM_SAMUEL_HALL_MEASURE_REPORT%202017.pdf [Accessed 15 February 2022]
- Saris, W.E. & Revilla, M. (2015) Correction for measurement errors in survey research: necessary and possible. *Social Indicators Research*, 127(3), 1005–1020.
- Schmier, J.K. & Halpern, M.T. (2004) Patient recall and recall bias of health state and health status. *Expert Review of Pharmacoeconomics & Outcomes Research*, 4(2), 159–163.
- Scott, J. & Alwin, D. (1998) Retrospective versus prospective measurement of life histories in longitudinal research. In: Giele, J.Z.Z. & Elder, G.H. (Eds.) *Methods of life course research: qualitative and quantitative approaches*. Thousand Oaks, CA: Sage Publications, pp. 98–127.
- Shachar, R. & Eckstein, Z. (2007) Correcting for bias in retrospective data. *Journal of Applied Econometrics*, 22, 657–675.
- Smith, J. & Thomas, D. (2003) Remembrances of things past: test–retest reliability of retrospective migration histories. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 166, 23–49.
- Smith, J.A. & Todd, P.E. (2001) Reconciling conflicting evidence on the performance of propensity-score matching methods. *American Economic Review*, 91(2), 112–118.
- Solga, H. (2001) Longitudinal surveys and the study of occupational mobility: panel and retrospective design in comparison. *Quality and Quantity*, 35(3), 291–309.
- SORA. (2020) *Sars-cov-2 in Austria. PCR tests in a representative sample*. Study report. SORA Institute for Social Research and Consulting. Vienna.
- Teitler, J.O., Reichman, N.E. & Koball, H. (2006) Contemporaneous versus retrospective reports of cohabitation in the fragile families survey. *Journal of Marriage and Family*, 68(2), 469–477.
- Thigpen, C. (2019) Measurement validity of retrospective survey questions of bicycling use, attitude, and skill. *Transportation Research Part F: Traffic Psychology and Behaviour*, 60, 453–461.
- Wang, J. (2021) To use or not to use propensity score matching? *Pharmaceutical Statistics*, 20(1), 15–24.
- Yan, T. & Tourangeau, R. (2007) Fast times and easy questions: the effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, 22(1), 51–68.
- Yarrow, M.R., Campbell, J.D. & Burton, R.V. (1970) Recollections of childhood a study of the retrospective method. *Monographs of the Society for Research in Child Development*, 35(5), iii.
- Zhou, X. (2000) Economic transformation and income inequality in urban China: evidence from panel data. *American Journal of Sociology*, 105, 1135–1174.
- Zurbriggen, C.L.A., Jendryczko, D. & Nussbeck, F.W. (2021) Rosy or blue? Change in recall bias of students' affective experiences during early adolescence. *Emotion*, 21(8), 1637–1649. Available from: <https://doi.org/10.1037/emo0001031>

How to cite this article: Dennison, J. (2023) The reliability of recall measurement in assessing migrant reintegration: Evidence from Ethiopia, Somalia, and Sudan. *International Migration*, 00, 1–18. Available from: <https://doi.org/10.1111/imig.13177>