

IS THERE AN EVIDENCE BASIS FOR IMMIGRANT INTEGRATION POLICIES? *A Methodological Enquiry*

Abstract

More than 10 years after the introduction of immigrant integration policies in Western Europe, policy debates have shifted from conceptual issues (What is integration? When is an immigrant integrated into a society?) to the effectiveness of integration policies (Do integration policies help immigrants to find jobs, learn a language or access services?). There is, in other words, a manifest need of an evidence basis, but no scientific studies on the topic have been published so far. Is there an evidence basis for immigrant integration policies? Which standards ought such an evidence basis to fulfil?

This article addresses the two questions from an exploratory perspective. It provides an overview of the available information on the effectiveness of immigrant integration policies across the European Union and the Organisation for Economic Co-operation and Development. By doing so, it contributes to the incorporation of the perspective of evidence-based policy literature from the field of public administration into immigrant integration studies.

Keywords

Integration policies • Civic integration policies • Policy evaluation • Effectiveness • Evidence-based policy

Received 7 June 2015; Accepted 17 October 2017

1 Introduction

More than 10 years after the introduction of integration policies, the focus of policy discussions is shifting from the definition of integration to the effectiveness of policies and the quest for evidence of that effectiveness. This article deals with the following questions: Is there an evidence basis for integration policies? Which conceptual and methodological standards should such an evidence basis display?

Immigrant integration, defined as a 'a *dynamic, two-way process of mutual accommodation* by all immigrants and residents of Member States' (Council of the European Union 2004) fully became a policy priority in Europe only during the early 2000s with the introduction of several measures such as civic integration courses, language education and labour market policies directed specifically at newly arrived immigrants (Joppke 2007).

During the first years of the policies, the discussion focussed on how integration should be defined and measured. However, as the policies reach maturity, the focus has moved from the 'what' question to the 'what works' question: Do integration policies enhance immigrants' employment opportunities? Do they counter over-qualification? Do they help against wage discrimination? The question whether those policies 'work' or 'are effective', i.e. whether they do contribute to the accomplishment of the accommodation between immigrants and residents, has a central role in the national

Montserrat González Garibay¹, Peter De Cuyper^{2*}

¹University of Ljubljana, Slovenia

²HIVA Research Institute for Work and Society, KU Leuven, Belgium

and international policy discourse (Niessen & Huddleston 2009). At the European Union (EU) level, the concepts 'evidence-based policy' and 'policy evaluation' have been emphasised by the Common Basic Principles: 'Developing clear goals, indicators and evaluation mechanisms are necessary to adjust policy, evaluate progress on integration and to make the exchange of information more effective' (Council of the European Union 2004).

One of the actions needed to achieve the said goal is to provide a 'broader evidence basis for integration policies through research' (European Commission 2005). Such an evidence basis should tell us, in other words, whether and in which circumstances integration policies are effective. However, the official EU documents do not elaborate further on the characteristics that such an evidence basis should have, nor which standards the research should fulfil to be considered valid evidence.

Within this context, this article deals with those two gaps by posing the following questions: Is there an evidence basis for integration policies across the EU and OECD member states that have implemented them? Which conceptual and methodological standards should such an evidence basis for integration policies comply with? In other words, if there is an evidence basis, does it fulfil certain quality criteria? If there is none, which quality criteria should it fulfil? The article does not attempt to actually answer the question whether policies work or not by discussing the findings of different

* E-mail: peter.decuyp@kuleuven.be

studies, but it rather explores the methodological features of those studies in order to answer the two research questions.

So far, the construction of an evidence basis for integration policies as such has remained largely off the radar of the integration literature, which focusses on the conceptualisation of the integration concept behind the policies (e.g. Joppke 2007) and on what the policies' goals are or should be (i.e. when is an immigrant considered integrated?) (Freeman 2004), rather than on the results of the policies themselves. In this context, this article's relevance vis-à-vis the current academic literature on immigrant integration is twofold. First, it provides an exploratory overview of the existing sources of evidence on the effectiveness of integration policies. Second, it tests the existing evidence on integration against the evidence-based policy (EBP) literature from the field of public administration and its long tradition of assessment of policy effectiveness.

The review encompasses both the existing compilations of best practices and inventories at the international level of the EU, the Organisation for Economic Cooperation and Development (OECD) and the International Organisation for Migration (IOM), as well as 47 evaluations at the national and sub-national levels, collected by means of an explorative methodology. Our analysis attempts to be exhaustive, provided that three structural constraints are taken into account. First, not all EU and OECD member states have enforced integration policies for newcomers. Second, not all countries with integration policies have conducted publicly available evaluations of their programmes. Third, evaluations are often only available in the local language (as in the case of some Norwegian studies). Hence, the research was limited to those studies available in the languages covered by the research team (English, Dutch, Spanish, French, Italian and Portuguese). Studies for which English summaries were available were also incorporated in the analysis. An overview of the final selection of studies and their discussion is provided elsewhere (González Garibay & De Cuyper 2013).

The remainder of the article is organised as follows. The second section briefly illuminates the EBP concept and assesses the strengths and weaknesses of the different methods used for the construction of an evidence basis. The third section documents the existing evidence regarding integration policies across the EU and the OECD and assesses its methodological quality by considering the EBP literature in order to answer both research questions. Building on that assessment, the fourth section provides a blueprint for the construction of an evidence basis. The last section presents a summary and highlights the main lessons learned.

2 What Is an Evidence Basis?

2.1 Definition of an Evidence Basis

The concept of EBP arose during the 1970s and 1980s in the context of the New Public Management (NPM) administrative doctrine, which was aimed at improving the performance of public administration by introducing measurements, standards and output control on the basis of several criteria, among which was effectiveness (Hood 1991). In this context, 'effectiveness' points to the extent to which a policy achieves the goals for which it was created. To study the effectiveness of a policy is to answer the question: 'Does the policy work?'

The 'What works?' question and the related 'What works where, for whom and in which circumstances?' form the basis for EBP, which points to the connection between scientific knowledge and

policy-making. EBP is defined by Davies, Newcomer and Soydan (2006: 175) as an approach that 'helps people make well-informed decisions about policies, programs and projects by putting the best available evidence from research at the heart of policy development and implementation'. In other words, the evidence-based approach is related to the use of 'objectivised' information from empirical evidence and evaluation of policy-making, i.e. information that is as free as possible of ideology and that has been obtained by applying scientific methods, as opposed to pure ideology or political rationality (for a discussion of the concept, refer Shaw *et al.* 2006: 489–501).

Answering the 'what works' question assumes that two needs are met. First is the need to gather data about the policies, i.e. policy monitoring (Niessen & Huddleston 2009). Monitoring implies gathering information about policy results, in terms of both outputs and outcomes. Outputs are a policy's direct results, such as the number of language courses that newcomers attend. However, most policies go beyond outputs and aim at producing effects (outcomes) such as employment. Whereas outputs lie within policymakers' direct sphere of control, outcomes are often behavioural changes beyond their control.

Second, answering the 'what works' question requires the establishment of a causal link between a policy (its inputs and outputs) and its outcomes. The question of causality goes back to David Hume (2007) and is characterised by several opposed traditions. Within the EBP literature, Pawson (2008) has framed the causality debate along a threefold distinction. In successionist causality, variables are associated with certain results by means of experiments or surveys in which a control group is tested. Configurationist causality looks at combinations of attributes of cases that explain variations. Generativist causality seeks to identify the mechanisms (i.e. processes describing human actions) that lead to certain outcomes.

2.2 Methodologies for the Construction of an Evidence Basis

The methods used for the construction of an evidence basis depend strongly on the view on causality that is adopted. Petticrew and Roberts (2007) mention several types of literature reviews by which research is condensed: systematic, narrative, conceptual, rapid, realist, scoping, traditional, critical, expert and 'state of the art'. Here, some examples representative of the three causal views are provided: the systematic review, the realist review or synthesis, and the effectiveness ladder.

A systematic review is 'a review of a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant research (...)' (Schibel *et al.* 2002: 2). The systematic review features a successionist perspective, in which causality (and thus effectiveness) should ideally be approached with experiments (Head 2008). The methodology was introduced in the social sciences¹ in order to compile scientific knowledge about policy effectiveness. Regarding the concrete way in which systematic reviews are conducted, Petrosino *et al.* (2001 cited in van der Knaap *et al.* 2008: 51-52) set out five steps to conduct a systematic review: the formulation of research questions; the determination of inclusion and exclusion criteria; the search for potential studies; the screening of studies for eligibility; and the analysis, synthesis and presentation of the selected studies. Weights are assigned to the different types of quantitative methods used by the studies in question, which range from correlations to experimental methods. Those reviews are conducted, distributed and inventoried by well-established networks

such as the Cochrane Collaboration (healthcare) or the Campbell Collaboration (social policies). Governmental instances have made use of the reviews in order to base their policy decisions (see for instance Department of International Development 2013).

In spite of the systematic review's apparent popularity, several criticisms have been voiced, which are mainly related to the positivist perspective on causality. Systematic reviews only include studies that explore more or less linear relationships between policy interventions and effects, while leaving configurational and generative causal patterns unexplored (Pawson & Tilley 1994), so we know whether a policy works, but not *why, for whom and in which circumstances*: the implementation of policies remains a 'black box' in which the exact mechanisms by which social reality interacts with elements of the policies in question remain obscure. This may lead to overlooking of the *undesired* effects of a policy.

In view of the critiques, several authors have attempted to offer alternatives. Dixon-Woods *et al.* (2006) argue for instance that qualitative information should be included in systematic reviews. Building on generative/configurationalist causality (Pawson & Tilley 1994; van der Knaap *et al.* 2008), some researchers have launched the concept 'realist synthesis' (Pawson *et al.* 2004), which focusses on the mechanisms linking policies to outcomes. The realist review starts with the proposition of a theory of change that outlines the causal mechanisms being active within the policy intervention at stake. That theory guides the collection of information sources.

The realist synthesis, while compensating for some of systematic reviews' weaknesses, displays some limitations: it emphasises policy description above effectiveness, it focusses on single policy instruments as opposed to broad questions, and it requires information on informal decision-making (Pawson *et al.* 2004: 11).

A third concrete proposal, the effectiveness ladder, originating from the domain of social work (Steyaert 2010; Veerman & van Yperen 2007), attempts to combine the systematic review's focus on effectiveness with the realist review's qualitative focus. To this end, Veerman and van Yperen (2007) propose a classification and quotation system for policy evaluations, called the 'effectiveness ladder'. In this system, both quantitative and qualitative studies can be incorporated within an evidence basis, and these receive a certain weight according to their methodological quality. That methodological quality is defined in function of the extent to which they prove, using statistical methods, causal links between a policy intervention and its effects (i.e. effectiveness). Experimental methods have the largest weight, as they offer methodological certainty. Descriptive studies have the smallest weight, as they do not offer any proof of their claims. The ladder is summarised below in Figure 1.

The first level contains descriptive evidence about policy interventions (goals, target groups and techniques), from which conclusions about the potential effectiveness of measures can be drawn. The second level includes evidence on the plausible effectiveness ('why and how intervention activities with a particular target group will lead to the intended outcomes'). The third level refers to indicative evidence, which shows that the effects aimed at by a certain policy do indeed occur but using methods that offer less certainty than experiments, such as monitoring studies. The fourth level includes causal evidence on the basis of randomised controlled trials and repeated case studies in which a single individual is intensively followed up.

The ladder is meant to be, rather than a scientific concept, a guidance tool for policymakers when choosing instruments. For instance, a database based on the ladder has been created in the Netherlands in order to classify youth policy interventions (Netherlands

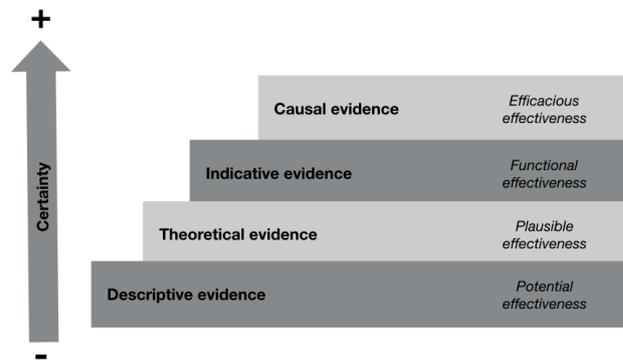


Figure 1. The effectiveness ladder
Source: Veerman and van Yperen (2007: 216, own modifications)

Jeugdinstituut 2017). Public and private actors can obtain information about whether a certain policy instrument has worked or not, as well as the circumstances in which effects have taken place.

Whereas the use of the effectiveness ladder, as opposed to the systematic review, implies less methodological certainty about policy effectiveness, it may provide information about its context and causal processes. The effectiveness ladder may nevertheless entail some pitfalls. Steyaert (2010) points to the fact that reviews using the ladder may remain biased towards experimental evidence, which remains the highest methodological standard. Moreover, we may ask the question about the role of descriptive evidence: does it ultimately offer any evidence about *why* or *how* a policy works?

3 Is There an Evidence Basis for Integration Policies?

3.1 Methodology

It is clear from the previous paragraphs that the concept 'evidence basis' is linked at its core with the causal link between a policy intervention and its effects. In the light of both research questions, we may ask the following questions: a) Have causal links been established in the field of immigrant integration (i.e. is there an evidence basis?) and b) If so, which features do they display or should display?

In order to answer the question, several sources were compiled following an exploratory methodological strategy. The main aims of the compilation were to conduct an exhaustive analysis of the existing inventories of evidence, as well as to collect as many individual studies as possible that assessed the broadest possible variety of instruments and using the broadest possible variety of methods.

The strategy was fivefold. First, the concept 'effectiveness' was broadly defined in terms of successionalist, generative and configurational causality, as to include both quantitative and qualitative studies. Studies providing exclusively descriptive evidence were left out. Second, the integration policies for which evidence was to be sought were defined operationally in terms of the EU's definition of integration (cf. supra, Introduction). That definition was constrained by focussing the overview on policies that explicitly target the accommodation of first-generation immigrants. Third, the available websites and compilations of best practices produced by the EU, OECD and IOM (such as European Commission Directorate-General

for Migration and Home Affairs [DG Home] 2011; IOM 2010; OECD 2007) were assessed. Fourth, an internet search of national and sub-national evaluations, published in academic journals or elsewhere, was conducted, in which only studies describing the outputs or assessing the effectiveness of policies (i.e. the link between policies and outcomes), either by quantitative or by qualitative means, were retained. The search was conducted using Google and Google Scholar, by applying a snowball technique. All of the evaluations that were mentioned in the international compilations were looked up, and their bibliographies were scanned to find more relevant sources. Those publications quoting them were tracked down and chosen if relevant. In addition, several queries ('immigrant integration', 'civic integration', 'evaluation' AND 'integration' AND 'immigrants', 'evaluation' AND 'civic integration') were used to find additional articles. Thus, 47 studies (policy evaluations or scientific articles) assessing the outputs and effectiveness of integration policies were retained. For some of the studies referenced in the compilations, it was the information in the compilation that was used instead of the original study due to language or availability issues. Fifth, the 47 selected references were classified according to their methodological approach towards integration (cf. *infra*, Tables 1–3). The remainder of this section provides a description of both the different sources and their methodological approach.

3.2 The Academic Literature

The academic literature on integration and civic integration policies has been focussed so far on the *philosophy* and the *politics* of integration policies rather than on the *effects* of those policies: topics include citizenship (Goodman 2009, 2011; Joppke 2007), the politics and governance of civic integration (Jørgensen 2012; Poppelaars & Scholten 2008), the interactions between integration research and policy-making (Scholten 2007) and the way in which integration policies should be judged and compared (Niessen & Huddleston 2009). In other words, evaluations of integration policies within the field of integration studies are almost non-existent². The few academic articles addressing the effects of integration policies are situated within the field of economics³. In that area, only Rinne (2012) provides an overview of 18 quantitative evaluations of integration policies. He includes several types of instruments: immigrant selection and settlement policies, introduction programmes, labour market policies, language courses and anti-discrimination policies.

3.3 International Compilations

So far, most of the attention to the evaluation of integration policies has come from the policy field rather than the scientific literature. The move towards the construction of an evidence basis at the policy level started during the 1990s with the collection of indicators on integration (Niessen & Huddleston 2009: 19–21). Following the introduction of the Common Basic Principles on immigrant integration, several initiatives were adopted to promote policy learning.

First, collections of policy measures adopted by the EU member states have been produced in the form of a website (European Commission 2017) and a series of handbooks on integration. The handbooks include policy recommendations based on successful practices and lessons learned.

Second, the Draft European modules on migrant integration attempt to go a step further than the Handbook by 'providing

Member States with negotiated recommendations on how to improve their integration policies and practices, based on the best existing evidence of what works' (European Commission. DG Home 2011: 3). The modules are a conscious attempt to build an evidence base for integration policies. Drawing information from the Handbook, they look for best practices and classify them along a fourfold scheme: 'strong evidence', ('an evaluation/research [...] shows that the [...] measure has a positive impact'), medium evidence ('participant or user evaluation' or monitoring suggest a positive impact), low evidence (best practices identified by experts) and best practices on how to do monitoring and evaluation.

Third, the European Integration Fund, which provides project funding, has established an evaluation requirement for the projects it finances. The Fund is also a contributor to the Migrant Integration Policy Index (MIPEX), a tool developed by an independent think-tank on the basis of expert interviews, to compare the openness of integration policies towards newcomers. The index scales integration policies in several domains: education, family reunion, residence, labour market mobility, access to nationality, political participation and anti-discrimination. Fourth, work has been conducted on the production of indicators that both grasp the situation of third-country nationals within the EU in several domains such as work, education or health, and that allow to compare countries with each other (Eurostat 2011). The most recent report on the use of integration indicators (Huddleston, Niessen & Tjaden 2013) mentions EBP ('multivariate and longitudinal analysis' and 'econometric and causal evaluations of policy impact') as one of the possible applications of the indicators.

Outside the EU institutions, comprehensive studies about the labour market integration of immigrants have been produced by the OECD and the IOM. The IOM edited a compilation of labour market integration policies. The OECD has assessed the integration of immigrants in 11 of its member states (OECD 2007, 2008, 2012). Its studies provide a comprehensive description of policy instruments, indicators on labour market integration and information on policy effectiveness (where available), based on own analyses and existing sources.

3.4 Evaluations at the National and Sub-National Levels

At the national and sub-national levels, several countries have produced evaluations of their integration and civic integration policies. These evaluations are characterised by a strong variation along many parameters: the type of policies being evaluated, the modalities of the policies, the purpose of the study (explanatory vs. descriptive), the data collection methods (administrative databases, surveys and interviews) and the statistical validity of the methods that were used (quasi experiments, longitudinal studies, etc.). This variety translates into large differences regarding the methodological quality of the evaluations. The remainder of this section focusses on each of the dimensions.

As said earlier, the policies under evaluation in the retained studies encompass several policy instruments: 12 evaluations deal with civic integration policies as a whole (these policies most often include cultural orientation, possibly along language or career guidance), nine focus on language courses, 10 on labour market policies (career guidance, subsidised employment, internships and training), four on mentoring (i.e. matching natives and immigrants for professional or social goals), two on pre-departure orientation (i.e. provision of information prior to arrival in the host country), three on refugee dispersal (i.e. obligatory settlement of refugees

in certain areas of the host country) and seven on support services (one-stop shops, general information provisions). The domains they emphasise vary strongly along geographical lines: evaluations of labour market policies are mainly concentrated in Northern European countries, whereas Central–Western European countries (Belgium, the Netherlands, Germany) also pay attention to cultural aspects.

Second, the policies differ strongly in their modalities: they can be either compulsory or voluntary, require the payment of fees or be free of charge, be implemented by governmental agencies or charities or may depend on national, regional and local governments.

Third, the purpose of the evaluations displays a strong variation. In an evaluation context, descriptive studies can be associated, in the first instance, with the description of policy outputs and processes (i.e. policy monitoring), whereas explanatory studies look for the causal link between interventions and effects (i.e. effectiveness). As can be seen from Table 1, a large majority of the studies retained for the overview attempt to go beyond the purely descriptive level to explain the dynamics behind the interventions they evaluate. This is valid across all types of integration policies.

This does not imply, however, that causal evidence is provided by all studies aiming at conducting effect evaluations. This depends, among others, on the type of data used in the study and on the methods by which causality is dealt with.

A fourth criterion regards the type of data used for the analysis. As already stated, both descriptive and explanatory goals can be pursued by using several types of data collected by several means: quantitative (administrative databases or survey) or qualitative (document analysis, interviews and focus groups). Table 2 provides an overview of the types of data used by the studies. Note that as one study may use more than one method, the totals do not add up to 47.

It is clear from the table that survey and administrative data are used to a considerably larger extent than qualitative data, with the exception of support services and mentoring. Evaluations of civic integration policies are more or less evenly spread across quantitative and qualitative data. It should be noted, however, that ‘administrative data’ may refer to both large administrative datasets involving strict methodological requirements and containing information for large parts of a country’s population, as well as administrative data referring to a single intervention, collected directly from participants. In other words, the methodological quality of the study cannot be fully assessed by only looking at the data.

Table 3 offers some complementary information about the methods used. It should be noted that several studies use a variety of methods. When classifying the studies, a priority rule was defined in which longitudinal methods have the largest importance, followed by cross-sectional regressions, descriptive statistics and qualitative analysis.

As opposed to the previous tables, this one displays a marked difference across types of policy interventions. Civic integration policies are mainly evaluated by means of descriptive statistics, whereas language courses are the subject of cross-sectional and longitudinal studies in which causal methods (i.e. regressions) are applied. Labour market policies are mainly evaluated by longitudinal studies in which duration models are applied to complex administrative datasets, as well as quasi-experimental settings with control groups. These methods are applied to all evaluations concerning refugee dispersal policies as well.

Table 1. Purpose of the selected evaluations

	Descriptive	Explanatory	Descriptive and explanatory	Total
Civic integration policies	4	2	6	12
Language courses	2	1	6	9
Labour market policies		9	1	10
Mentoring		4		4
Pre-departure orientation programmes			2	2
Support services		3	4	7
Refugee dispersal		3		3
Total	6	22	19	47

Source: González Garibay and De Cuyper (2013)

Table 2. Types of data used by the selected evaluations

	Quantitative data		Qualitative data	Unknown
	Administrative	Survey		
Civic integration policies	5	4	7	1
Language courses	9	2	1	
Labour market policies	6	5		
Mentoring	1	1	3	1
Pre-departure orientation programmes		2	2	
Support services	3		6	
Refugee dispersal	3			
Total number of times a type of data is used	27	14	19	2

Table 3. Methods of analysis used by the selected evaluations

	Quantitative (unknown)	Qualitative analysis	Descriptive statistics	Cross-sectional regressions	Longitudinal data analyses (logistic regressions)	Total
Civic integration policies	2	1	9			12
Language courses		1	2	3	3	9
Mentoring	1	2	1			4
Pre-departure orientation programmes			2			2
Support services			4	3		7
Refugee dispersal					3	3
Total	3	8	18	6	12	47

3.5 Is the Available Evidence an Evidence Basis?

The policy and academic sources mentioned earlier offer information on the situation of immigrants across the EU and the OECD, as well as on the policies implemented by the member states. However, the question arises whether they constitute a basis of evidence.

As stated earlier, the central notion within the evidence-based approach is the focus on impact or effectiveness (causality). When contrasted with that definition, both the inventories of integration policies and the individual studies described herein display several shortcomings. First, the EU sources do not address the effectiveness of policies explicitly and thus do not set forward a concrete vision of causality: ultimately speaking, we do not know along which causal lines they define 'impact'. Even though large amounts of information on both the situation of immigrants and on the policies directed at influencing that situation have been collected, the exact way in which the policies relate to that situation remains obscure. Even though the Draft Modules address the construction of an evidence basis, they cannot strictly be regarded as a full-fledged evidence basis in the sense of the systematic review, the realist synthesis or the effectiveness ladder, as they do not define causality nor take methodological considerations into account to classify information.

Moreover, although MIPEX constitutes an attempt to bridge the gap between the indicators and the policies, it does not measure the *actual* effectiveness of integration policies but rather judges their *potential* effectiveness on the basis of experts' opinions.

Second, the methodology to select the EU's best practices is not clear: the Handbooks on integration do not offer any information on how the good practices were selected, nor do they systematically provide information on the results of those best practices. In other words, we lack information about the way in which the concept 'best practice' is defined. The Draft Modules, which use the Handbooks as the main source, display the same shortcoming.

Third, even though the OECD studies do address the effectiveness of integration policies, their assessment displays two shortcomings. The studies focus on the effectiveness of policies with respect to labour market integration, leaving aside the cultural aspects of the integration process, which involve for instance networking with the local population and feelings of identification with the receiving society (Bijl & Verweij 2012). In addition, they only assess the effectiveness of certain types of instruments (language courses, labour market measures and settlement policies) for which quantitative evidence

exists, leaving aside instruments such as mentoring and pre-departure orientation. The OECD studies suffer in other words from a limitation regarding the way in which the integration concept is operationalised (it remains limited to the labour market) and to the types of instruments it studies. Moreover, the OECD reviews do not provide any information on the methodological quality of the studies on which they are based: even though most of the studies assessed use quantitative methodologies, we know little about the quality of those methods and the information they used. Rinne's (2012) study mentioned earlier shares the same features.

Fourth, the fact that the national studies have not been compiled and synthesised makes it difficult to compare their methodological quality and the results they present.

A last consideration refers to the methodological quality: in spite of the fact that a large majority of the studies (n=41) aim at explaining policy outcomes, only a minority (n=18) uses methods by which some certainty regarding causality patterns can be obtained. In addition, all of the reports published outside academic journals offer little information on methodological considerations. This makes it difficult to grasp the vision of causality held by the authors.

The consequence of the gaps is that the available academic and policy sources on the effectiveness of immigrant integration civic integration policies remain scattered. This lack of knowledge has already been remarked by Niessen and Huddleston (2009), who plead for an evaluation chain on integration policy in which the effectiveness, efficiency and sustainability of policies are measured through the collection of a number of indicators. In short, there seems to be a large amount of evidence, but it has not been fully integrated into an evidence basis constructed along strict methodological and conceptual principles.

When we look beyond the compilations and the academic literature as a whole and into the individual studies referred to in such compilations, a further concern arises: regardless of whether the compilations themselves form an evidence basis, the methodological quality of the available evidence on integration policy may not be enough to comply with the requirements of a 'classic' evidence base in the form of a systematic review. Even though a large majority of the studies state the intention to assess policy effectiveness, the proportion among them using methods by which that effectiveness can be assessed with statistical validity (i.e. cross-sectional or longitudinal studies using regressions) is reduced, as well as being concentrated in a very limited number of countries. There is, in other

words, no 'critical mass' of statistically significant studies. This has some implications for the further construction of an evidence basis, which are highlighted in the following paragraphs.

4 Which Method to Use for the Construction of an Evidence Basis for Integration Policies?

Considering our first research question (is there an evidence basis?), the answer is negative: the previous section has made it clear that the existing evidence for the effectiveness of integration policies has not been integrated so far in an evidence basis. The second research question arises in this context, i.e.: What is the adequate type of evidence basis: a systematic review, a realist synthesis or an effectiveness ladder?

There are a number of features of the existing evidence presented herein that point to the effectiveness ladder. First, the large variety of instruments and the limited number of studies make it necessary to create an inclusive evidence basis that encompasses both quantitative and qualitative research. Second, given the methodological variety of the instruments and the limited number of quantitative studies that would qualify for a systematic review ($n=19$), such a technique would only offer a limited insight into policy effectiveness and leave out qualitative research. This echoes the findings of Schibel *et al.* (2002), who conclude that a systematic review of the field of refugee policies would leave out a considerable number of evaluations, given its strict criteria. Third, given the fact that integration and civic integration policies are fairly recent, there is little insight on *how* the policies have an impact on immigrant integration, which is explored in several of the qualitative studies. A systematic review, which focusses on *whether* policies work, would leave such studies out. In this context, it seems most adequate to choose the effectiveness ladder as the starting point for the construction of an evidence basis. This does not imply, however, that the systematic review should be excluded in the long term. As integration policies become more widespread and the use of administrative databases increases across the EU, it is likely that more quantitative, quasi-experimental studies are carried out about these policies.

As already pointed out, the effectiveness ladder was developed with the specific characteristics of youth care in mind. The question then arises here whether the ladder should be adopted in its entirety: Which types of measures should be included in the evidence basis? Which kinds of methods should be included? Can we open up Veerman and van Yperen's (2007) classification to encompass integration policies?

A first consideration when answering this question refers to the *scope* of the types of measures to be included in such a review. As argued earlier, the EU definition is only a lowest common denominator regarding the definition of integration across countries, and at the same time, a large variety of policy instruments are used in the field. The question arises whether all those policies may be included in a single review or a single evidence basis. In this context, we plead to limit the scope of the measures included in an evidence basis to the smallest common denominator of OECD countries' integration policies: they are often directed at recent immigrants, whichever definition of immigrant they provide. In this context, the evaluations we suggest to include refer to policies explicitly aimed at the integration of persons of foreign nationality, or the integration of adults of foreign origins who reside in the country applying the policy under any type of scheme (refugees and asylum-seekers, family migrants and labour migrants) or who have received the authorisation to establish

themselves in that country. They need not be the only target group: other groups such as second-generation immigrants or travellers may also be involved in the policies, as well as the receiving society⁴ (e.g. mentoring schemes). However, they need to be a specific target for the measure to be included in the evidence basis.

A second consideration refers to the methods taken into account by Veerman and van Yperen. Whereas most of those methods have also been applied in the evaluation of integration policies, there are some exceptions. The causal methods they mention include randomised controlled trials and repeated case studies ($n=1$ designs), in which one client is 'carefully observed before treatment and from the moment that treatment starts'. Such studies have not been extensively conducted in the field of integration. Moreover, the authors group methods that provide different types of causal evidence in a single category (indicative evidence). On the one hand, client satisfaction, goal attainment, monitoring and quality assurance studies yield some evidence about the effectiveness of a measure; yet the causal relationship between the intervention and its effects may or may not be directly established.

In order to take the above remarks into consideration, we propose a new effectiveness ladder in which a) a clear policy scope is defined and b) account is taken of the different types of (semi-)causal evidence delivered by different indicative methods. The proposal is summarised in Table 4.

This effectiveness ladder may be regarded as a first blueprint for the construction of an evidence basis that focusses on the effectiveness of integration policies while preserving a broad concept of effectiveness and leaving enough room for alternative visions of causality.

5 Concluding Remarks

This article started with two questions: Is there an evidence basis for integration policies across the OECD countries? Which conceptual and methodological standards should such an evidence basis comply with?

In order to answer the questions, the article first dug into the EBP literature to shed some light on the evidence basis concept and illustrated the standards by which an evidence basis is made by means of three examples: the systematic review, the realist synthesis and the effectiveness ladder. Those evidence models may have a specific vision of causality: whereas the systematic review leans on successionist causality, the realist synthesis centres on generative causality (mechanisms) and the effectiveness ladder still looks at successionist causality as the main standard for defining the effectiveness of a policy while allowing other types of causality in the evidence basis. Each vision has specific consequences for the standards along which the evidence basis should be constructed: systematic reviews rely on experiments, whereas realist syntheses and the effectiveness ladder allow for both quantitative and qualitative sources.

On the basis of the literature and the three sorts of evidence bases, the article contrasted the existing evidence on integration policies against that literature. The existing evidence was compiled along a twofold strategy that aimed at being as inclusive as possible. On the one hand, the initiatives undertaken by the EU, the OECD and the IOM to compile the available evidence of the policies on immigrant integration were assessed. On the other hand, 47 studies of integration policies, either included in those compilations or published by governments or in academic journals, were compiled and assessed.

Table 4. An effectiveness ladder for integration policies

Levels of evidence	Parameters of evidence	Methods	Effectiveness	Scope of the research
5. Causal	As in 1-4, but there is sound and substantial statistical evidence that the outcome is caused by the intervention and/or which ingredients are responsible for the outcome.	Randomised controlled trial	Efficacious	Policy instruments with immigrants as direct target group
4. Semi-causal	As in 1-3, but there is evidence that the outcome is caused by the intervention and/or which ingredients are responsible for the outcome. However, alternative explanations cannot be discarded with certainty	Quasi-experimental studies, theories of change studies, norm references approaches	Likely efficacious	
3. Indicative	As in 1 and 2, but now there are indications	Client satisfaction studies, goal attainment studies, monitoring studies, quality assurance studies	Functional	
2. Theoretical	As in 1, but intervention now has a possible rationale to explain why it should work, with whom	Reviews, meta-analyses, expert knowledge studies	Plausible	
1. Descriptive	The essential elements of the intervention (e.g. goals, target group, methods and activities, requirements) have been made explicit	Descriptive studies, observational studies, analysis of documents, conduct of interviews	Potential	

Source: Veerman and van Yperen (2007, own modifications).

The outcome of the exploratory exercise can be summarised in four main lessons learned. First, in the light of the first research question, the existing sources of evidence regarding the effectiveness of integration policies cannot be considered a full-fledged evidence basis. The compilations that have been produced at the international level lack a vision of causality. In other words, they do not tell us what exactly is considered an effective policy. Moreover, the way in which they classify individual evaluation studies or best practices remains murky, as they do not develop clear methodological standards.

Second, the 47 studies that were compiled and analysed display a very broad variety along several criteria, such as the type of instruments assessed, the implementation modalities of those instruments and the way in which they view causality and the effectiveness concept.

Third, in the light of the second research question, if we are to bring some structure into that variety and set some steps towards the construction of an evidence basis, there are two basic standards to be followed. On the one hand, an evidence basis for integration policies needs to choose a specific perspective or vision on effectiveness, i.e. the causal relationship between a policy and its outcomes. On the other hand, it needs a clear categorisation of the type of studies or information sources that will be used as evidence.

Fourth, taking into consideration these two standards, the effectiveness ladder seems to be the most adequate method, at this stage, to start the construction of an evidence basis for integration policies. Integration policies are relatively young, as they only came into the EU mainstream scene during the second half of the 2000s. Their evaluation, which only started a few years after, has occurred along a variety of methods and approaches. Hence, it becomes necessary to grasp that a variety of methods and approaches exist in the evidence basis. In this sense, an effectiveness ladder offers a clear added value, since it allows us to systematise several sources and to identify gaps in the evaluations of certain instruments, such as the lack of process evaluations of labour market instruments

targeted at immigrants (cf. supra). In addition, it makes possible to include evaluations based on qualitative research focussing on processes and the *why* and *how* of integration policies, for which the knowledge seems limited so far. Conversely, the systematic review's methodological rigidity and the narrow focus of the realist synthesis on a single theory risk to reduce the scope of the evidence basis and fail to identify the research gaps. We should remain aware, however, of the fact that the effectiveness ladder still gives the most weight to experimental studies demonstrating successionist causal relationships and attempts to balance that weight by paying enough attention to other types of sources and, if needed, by adapting the weight given to those studies.

Montserrat González Garibay holds a PhD in Social Sciences (KU Leuven, 2011) and is currently pursuing a second PhD in Statistics at the University of Ljubljana. She has worked at the Research Institute for Work and Society (HIVA) and at the Flemish Ministry of Work and Social Economy, focussing on policy monitoring and evaluation.

Peter De Cuyper holds a Master's degree in Sociology (KU Leuven, 1998). He is research manager at the Research HIVA-KU Leuven. Currently, he focusses on the impact of labour market policies on immigrant employment, policies towards highly skilled newcomers and mentoring schemes for new migrants.

Notes

1. The systematic review methodology was first developed throughout the 1980s, within the domain of medical interventions (Dixon-Woods et al. 2006).
2. The effects of integration policies have been assessed to a limited extent by two projects: Integration of Third Country Nationals (INTI, see ICMPD 2005) and Integration and

Naturalisation tests: the new way to European Citizenship (INTEC, see Böcker & Strik 2011).

3. See for instance Joona and Nekby (2012).
4. This criterion excludes several policies that relate to integration but are targeted at the whole society, such as diversity-related advertising campaigns or anti-discrimination measures.

References

- Bijl, R & Verweij, A (eds.) 2012, Measuring and monitoring immigrant integration in Europe. Integration policies and monitoring efforts in 17 European countries, The Netherlands Institute for Social Research (SCP), The Hague.
- Böcker, A & Strik, T 2011, 'Language and knowledge tests for permanent residence rights: help or hindrance for integration?', *European Journal of Migration and Law*, vol. 13, no. 2, pp. 157-184.
- Council of the European Union 2004, *Press Release 14615/04 (presse 321)*, European Union, Brussels.
- Davies, P, Newcomer, K & Soydan, H 2006, 'Government as structural context for evaluation', in *The SAGE handbook of evaluation*, eds I Shaw, JC Greene & MM Mark, Sage, Thousand Oaks, pp. 163-183.
- Department of International Development 2013, *Systematic reviews in international development*. Available from: <<https://www.gov.uk/government/publications/systematic-reviews-in-international-development>>. [Last accessed 22.11.2016].
- Dixon-Woods, M, Bonas, S, Booth, A, Jones, DR, Miller, T, Sutton, AJ, et al. 2006, 'How can systematic reviews incorporate qualitative research? A critical perspective', *Qualitative Research*, vol. 6, no. 1, pp. 27-44.
- European Commission 2005, A common agenda for integration framework for the integration of third-country nationals in the European Union. COM (2009) 385, European Commission, Brussels.
- European Commission 2017, *European web site on integration*, Available from: <<https://ec.europa.eu/migrant-integration/home>>. [Last accessed 29.11.2017].
- European Commission. DG Home 2011, *Draft European modules on migrant integration*, European Commission, Brussels.
- Eurostat 2011, Indicators of immigrant integration. A pilot study, European Commission, Brussels.
- Freeman, GP 2004, 'Immigrant incorporation in western democracies', *International Migration Review*, vol. 38, no. 3, pp. 945-969.
- González Garibay, M & De Cuyper, P 2013, *The evaluation of integration policies across the OECD: a review*, Policy Research Centre on Integration, Antwerp.
- Goodman, SW 2009, Civic integration requirements and the transformation of citizenship, Georgetown University, Washington.
- Goodman, SW 2011, 'Controlling immigration through language and country knowledge requirements', *West European Politics*, vol. 34, no. 2, pp. 235-255.
- Head, BW 2008, 'Three lenses of evidence-based policy', *The Australian Journal of Public Administration*, vol. 67, no. 1, pp. 1-11.
- Hood, C 1991, 'A public management for all seasons?', *Public Administration*, vol. 69, no. 1, pp. 3-19.
- Huddleston, T, Niessen, J & Tjaden, JD 2013, Using EU indicators of immigrant integration. Final report for Directorate-General for Home Affairs, European Commission, Brussels.
- Hume, D 2007, *A treatise of human nature*, Clarendon, Oxford.
- ICMPD 2005, Integration agreements and voluntary measures. Comparison of compulsory integration courses, programmes and agreements and voluntary integration programmes and measures in Austria, France, Germany, the Netherlands and Switzerland, ICMPD, Vienna.
- IOM 2010, Migration, employment and labour market integration policies in the European Union. Part 2: Labour market integration policies in the European Union (2000-2009), IOM, Brussels.
- Joona, PA & Nekby, L 2012, 'Intensive coaching of new immigrants: an evaluation based on random program assignment', *The Scandinavian Journal of Economics*, vol. 114, no. 2, pp. 575-600.
- Joppke, C 2007, 'Beyond national models: Civic integration policies for immigrants in Western Europe', *West European Politics*, vol. 30, no. 1, pp. 1-22.
- Jørgensen, MB 2012, 'The diverging logics of integration policy making at national and city level', *International Migration Review*, vol. 46, no. 1, pp. 244-278.
- Nederlands Jeugdinstituut 2017, *Databank effectieve jeugdinterventies*, Available from: <https://www.nji.nl/nl/Databank/Databank-Effectieve-Jeugdinterventies>. [Last accessed 29.11.2017].
- Niessen, J & Huddleston, T (eds.) 2009, *Legal frameworks for the integration of third-country nationals*, Martinus Nijhoff Publishers, Leiden.
- OECD 2007, Jobs for immigrants. Volume 1. Labour market integration Australia, Denmark, Germany and Sweden, OECD Publishing, Paris.
- OECD 2008, Jobs for immigrants. Volume 2. Labour market integration in Belgium, France, the Netherlands and Portugal, OECD Publishing, Paris.
- OECD 2012, Jobs for immigrants. Volume 3. Labour market integration in Austria, Norway and Switzerland, OECD Publishing, Paris.
- Pawson, R 2008, *Causality for beginners. ESRC/NCRM Research Methods Festival*, Available from: <http://eprints.ncrm.ac.uk/245/>. [Last accessed 21.09.2013].
- Pawson, R, Greenalgh, T, Harvey, G & Walshe, K 2004, *Realist synthesis: an introduction*, University of Manchester, Manchester.
- Pawson, R & Tilley, N 1994, 'What works in evaluation research?', *The British Journal of Criminology*, vol. 34, no. 3, pp. 291-306.
- Petticrew, M & Roberts, H 2007, *Systematic reviews in the social sciences. A practical guide*, Blackwell Publishing, Malden.
- Poppelaars, C & Scholten, P 2008, 'Two worlds apart: the divergence of national and local immigrant integration policies in the Netherlands', *Administration & Society*, vol. 40, no. 4, pp. 335-357.
- Rinne, U 2012, The evaluation of immigration policies, IZA, Bonn.
- Schibel, Y, Fazel, M, Robb, R & Garner, P 2002, *Refugee integration: can research synthesis inform policy? Feasibility study report. RDS On-line Report 13/02*, Available from: <http://webarchive.org>.

- nationalarchives.gov.uk/20110218141325/http://rds.homeoffice.gov.uk/rds/onlinepubs1.html. [Last accessed 30.01.2018].
- Scholten, P 2007, Constructing immigrant policies. Research-policy relations and immigrant integration in the Netherlands (1970-2004), University of Twente, Twente.
- Shaw, I, Mowbray, CT & Qureshi, H 2006, 'Social work and the human services', in *The SAGE handbook of evaluation*, eds I Shaw, JC Greene & MM Mark, Sage, Thousand Oaks, pp. 486-511.
- Steyaert, J 2010, 'Where the worlds of e-inclusion and evidence-based practice meet', *Journal of Social Intervention: Theory and Practice*, vol. 19, no. 4, pp. 19-30.
- van der Knaap, LM, Leeuw, FL, Bogaerts, S & Nijssen, LTJ 2008, 'Combining Campbell standards and the realist evaluation approach: The best of two worlds?', *American Journal of Evaluation*, vol. 29, no. 1, pp. 48-57.
- Veerman, JW & van Yperen, TA 2007, 'Degrees of freedom and degrees of certainty: a developmental model for the establishment of evidence-based youth care', *Evaluation and Program Planning*, vol. 30, no. 2, pp. 212-221.