



German Council for Social  
and Economic Data (RatSWD)

[www.ratswd.de](http://www.ratswd.de)

# RatSWD

## *Working Paper Series*

Working Paper

No. 105

### Knowing More about Vocational Training

New Demands for Data and Research Infrastructure

---

Steffen Hillmert

---

July 2009

---

## Working Paper Series of the Council for Social and Economic Data (RatSWD)

---

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

*RatSWD Working Papers* are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/ 2008 Heike Solga; 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

# Knowing More about Vocational Training New Demands for Data and Research Infrastructure

**Steffen Hillmert**

*University of Tübingen (steffen.hillmert[at]uni-tuebingen.de)*

## **Abstract**

Modern societies depend on the successful and comprehensive provision of skills, and receiving vocational training in any form has been experienced by a majority of the population in younger cohorts. There has therefore been a constant demand for timely information about the various forms of training and their relations to a broader societal context. Over the recent decades, the patterns of participation in education and training have become more extended, more complex and more heterogeneous. Against this background, the paper discusses to what extent existing and projected data sources are suitable for investigating the relevant scientific and policy-related questions. Among these questions are: How does participation in training develop over the life course? What are the relative chances of receiving specific types of training; who, in particular, is likely to receive the most attractive types? Are training measures effective? When reviewing the current data situation, it becomes clear that progress has undoubtedly been made in the past few years. It is also obvious, however, that fundamental questions can presently not be answered on the basis of the available large-scale data on vocational education and training. Some key recommendations are presented.

Keywords: Vocational training; data; research infrastructure; overview; Germany

## 1. Analytical framework and research questions

Modern societies depend on the successful and comprehensive provision of skills, and receiving vocational training in any form has been experienced by a majority of the population in younger cohorts. From both a scientific and a policy-related perspective, there has therefore been a constant demand for timely information about the various forms of training and their relations to a broader societal context. This has raised questions like: How does participation in training presently look like? What are the relative chances of receiving specific types of training; who, in particular, is likely to receive the most attractive types? Are training measures effective?

Over the recent decades, the patterns of participation in education and training have become more complex and more heterogeneous. Young people do not necessarily receive training from one type of institution only, but they combine episodes of training in various ways (e.g., Hillmert and Jacob 2003). It therefore makes sense to speak of *careers* of education and training. Such careers have also become increasingly extended. Today many individuals participate in formal education for more than two decades. For research, this development emphasizes the necessity for studying education and training in a dynamic, life-course oriented perspective, which follows the requirements defined by empirical behavior, not administrative categories. There is considerable demand for detailed and reliable data on which educational policies can be based. While it is still important to know about aggregate numbers (like training contracts, participants or registered applicants at a particular point in time), it is especially necessary to look at individual situations and dynamic processes within education and training. This means that it is essential to collect not only “snapshot” information about current activities of vocational training, but also information about their connections with the previous life history and the following steps in a career.

In Germany, recent and ongoing studies on the basis of longitudinal survey data (e.g., German Socio-economic Panel; German Life History Study; BIJU; LAU/ULME; BIBB-Übergangsstudie; DJI-Übergangspanel) have collected valuable information about the educational careers – sometimes including measures of performance of competencies – and have thus enhanced our knowledge about likely causal relationships. Nevertheless evidence-based policies continuously require differentiated, reliable and up-to-date information. It is therefore necessary to raise also the quality of process-produced data according to the new demands. Official statistics are constructed on a specific legal basis, so any recommendation

of improvement has to account for the details of these regulations. Getting comprehensive data on the training situation of the whole population would require still more radical innovations in the information infrastructure like the introduction of a population register. Even when ambitions are smaller, there is an urgent need for improving the current data situation, and a short review reveals that even very basic information about the situation of vocational training in Germany is not available. Measures to change this would often require no more resources and yet be of great value. In any case, research will continue to be confronted with a variety of different data sources which should be coordinated as much as possible.

A simplified framework summarizes minimum requirements for the systematic collection of relevant information on training: It starts, first, from an *institution-oriented differentiation* as the basic condition for an assessment of the situation and the performance of the various parts of the vocational education and training (VET) system. Traditionally, vocational training statistics has often been equated with information on apprenticeships, but today it seems useful to distinguish between three major areas of the VET system below the tertiary level (Baethge et al. 2003; Autorengruppe Bildungsberichterstattung 2008): the *Dual System*, i.e. apprenticeships organized as a combination of firm-based training and vocational schooling; (full-time) *school-based training*; and also a large „*transition system*“ of measures including youth training schemes and basic forms of vocational training. There are also special cases like training in the Civil Service. Second, one can distinguish different *analytical elements* (or central “dependent variables” for empirical analyses) which represent the basic building blocks of careers in education and training. They allow making inferences about mobility and developments in the training system. These elements relate to questions of *access*, i.e. the transitions to specific forms of vocational training and their determinants; questions of progression and *outcome*, i.e. learning, competencies and qualifications in vocational training; and questions of *impact* in the sense of transitions out of vocational training and their consequences, in particular labor market consequences. Third, one needs to look at central dimensions of non-institutional and *individual-level differentiation* (as the set of central “independent variables” for the analyses) which allows comparing the relative chances of particular groups in society with regard to educational transitions and attainment. These analytical dimensions are summarized in Table 1. While variables like age and gender represent commonly used distinctions, differences with regard to nationality and – even more important for meaningful analyses – migration background have come only recently into the focus of official statistics. Another important form of non-institutional differentiation is

regional differentiation. Classifying various aggregate units may extend the data base further towards a multi-level structure.

It is essential for the analysis of training careers to dispose of longitudinal data. For research, *prospective* collection of information would be preferable, i.e. following the careers of individuals as they develop over time. However, such data designs have tended to be controversial (regarding issues of data protection) as they necessarily require matching information of individual cases across several waves of data collection. Hence, the second-best solution is collecting information on individual developments *retrospectively*, i.e. gathering time-referenced information about individuals' previous experiences and activities each time someone begins a (new) period of activities in VET.

Table 1: Systematic information for a dynamic analysis of vocational training

<b>Institution-oriented differentiation</b>	<b>Analytical elements in a life-course perspective</b>	<b>Individual-level and other non-institutional differentiation (examples)</b>
<ul style="list-style-type: none"> <li>- Dual System</li> <li>- School-based training</li> <li>- "Transition system"</li> </ul>	<ul style="list-style-type: none"> <li>- Access</li> <li>- Outcome/Performance</li> <li>- Impact/Consequences</li> </ul>	<ul style="list-style-type: none"> <li>- Individual history of education and training</li> <li>- Age</li> <li>- Gender</li> <li>- Migration background</li> <li>- Regional differentiation</li> </ul>

This framework is rather simple and selective. When contrasted with the basic questions in the first paragraph, however, it does make clear that fundamental and highly relevant questions (that have proved to yield significant findings in small-scale studies) can presently not be answered on the basis of the available large-scale data on VET. Among these questions are: How many people do combine two or more episodes of VET? Do young people with a migration background have similar chances of access to training than those without, how do they perform within the system? Does training within the dual system and in full-time vocational schools lead to similar patterns of transitions to employment? Do training programs result in transitions to regular training and/or stable employment? What proportion of a cohort entering vocational training does finally complete it successfully?

## 2. Contrasting demands and available data bases

This section discusses to what extent existing and projected data sources can be used to investigate such questions. Compared to other industrialized countries, VET in Germany is formalized to a relatively high degree. Due to the differentiation and the complex institutional structure of the German education and training system, however, the relevant information sources are very heterogeneous. A brief overview over relevant data and their characteristics – following the basic distinction between the three sectors of the VET system – is necessarily a selection, based on the centrality of the specific data sources (mainly official statistics), the timeliness of developments and the priority of necessary changes. For a general overview on the information infrastructure in this area, see Weishaupt and Fickermann 2000; Baethge et al. 2003, 47-54; Bellmann 2005; Brosi 2005).

### *(1) Data on vocational training within the Dual System*

Traditionally, the data situation regarding apprenticeship training has been comparatively good. Given its dual nature, it is covered by various data sources: the School Statistics (which are collected on the *Länder* level) and the Vocational Training Statistics of the ‘authorities in charge’, i.e. institutions like the Chambers of Commerce; these statistics are often referred to as *the Vocational Training Statistics*. There are also statistics on vocational counseling (e.g., the number of registered applicants and reported vacancies) which allow calculations of supply- and demand-relations on the training market, at least in a simple form (BMBF 2007). Additional information about apprenticeships is provided by sources like the Employment Statistics and a regular collection of data on firms (*IAB Betriebspanel*).

The Vocational Training Act BBiG (*Berufsbildungsgesetz*) is the central legal basis of the VET Statistics. As it is national law, it allows for a standardized gathering of information. The recent Reform Act has brought major improvements for the mandatory data collection (as of 2007), in particular the shift from aggregate reporting to individual-level information and the provision of more information on individual-level determinants of transitions (see Uhly 2006). Implementation of these changes is not yet finished. Still, there are considerable deficits for dynamic analyses: There is no linkage between years and no information on timing. In other words, there is information on transitions in a given year, but it represents a one-off snapshot. This means that there is insufficient information about the extension of training careers, the structure of multiple training episodes and other life-course aspects. Since termination rates are calculated on a yearly basis, it is not possible to distinguish between

temporary and final dropout and hence to calculate cohort-specific rates of success within the apprenticeship system.

*This situation could easily be remedied by the introduction of a personal ID number which would allow connecting individual-specific data entries across different years.* Currently, information on previous education and training – including school-based training and qualification measures – is collected (cf. Statistisches Bundesamt 2007a), but this information is restricted to type or level of these experiences and its range is limited. Matching person-specific data records across years (and institutions) is the preferable solution for reconstructing training biographies. As long as this cannot be done, it would be helpful to collect qualitative information on previous VET experience (most notably, occupation) and on the *timing* of previous experiences in training. Another specific deficit is insufficient information about a possible *migration background* which needs to go beyond information about the individual's nationality. It would be helpful to coordinate the definition of an extended set of collected variables with other central data sources in this area like the projected individual-level School Statistics (see the following section) where information may be available about country of birth, parents' country of birth, year of immigration (and language spoken at home). It may also be considered to include at least some proxy information on (previous) performance (in the form of grades). Proxy variables for regional mobility among apprentices may also be considered.

## *(2) Data on full-time school-based training*

The various forms of (full-time) school-based training have become an increasingly relevant part of the VET system. They form a mixed category, but a major share consists of training in occupations outside the BBiG/HwO regulations. Vocational schools are regulated by state law and covered by the School Statistics (Statistisches Bundesamt 2007b); data are collected by the *Länder* on the basis of different regulations and classifications. In many respects, the collected and the published information have been limited and heterogeneous (cf. already Krüger 2005).

A systematic deficit of statistics that focus on school students is that there is no information about applicants and (potentially) available places. This means that – in contrast to apprenticeships – there is no information about relations of supply and demand in this sector of VET. Special problems are associated with data on *school-based training in the healthcare sector*. Data is incomplete (e.g., there is no reporting in the federal state of Hesse)

and of varying quality. One of the reasons is that supervision lies with different authorities, some of them with no obligation to report.

The consequence of these characteristics is that there is still no comprehensive account of the volume and the structure of school-based training in Germany. Given these limitations and heterogeneity, major improvements can be expected from the coordinated ‘core data’ on schools recently projected by the *Länder (Kerndatensatz (KDS) für schulstatistische Individualdaten der Länder gemäß dem Beschluss der Kultusministerkonferenz)* as long as there will be comprehensive coverage of all vocational schools in all federal states. Similar to the changes in the BBiG statistics, this concept would also mean a shift from aggregate reporting to the collection of individual-level information and an extension of collected variables. For this concept to be significant for research on training careers, it is essential that individual records can indeed be matched from one year to another. However, implementation of this concept has raised some controversies and has so far proceeded very differently in the various federal states.

### *(3) Data on the “transition system” of training measures*

The term “transition system” subsumes a variety of measures in VET including youth training schemes and forms of preparatory vocational training (including attaining general school qualifications). It is a very heterogeneous category, but given its considerable size, it makes sense to include it in a concept of regularly monitoring VET. So far, it is not clear how transitory or permanent and how successful such measures are as parts of individual training careers. Empirical studies have relied on one-off samples or focused on special, temporary programs (Troltsch et al. 1999; Dietrich et al. 2002). For a specific sub-sample of the training population, there is analytical potential in matching different sources of process-produced data on (re-)training and employment as in the ongoing project “Integrated Employment Biographies” at the Institute of Employment Research (IAB).

Regular reporting about these measures is mainly based on the statistics of vocational schools and the statistics of measures financed by the *Bundesagentur für Arbeit*. It is difficult to sort out any overlapping in the reported figures (cf. Autorengruppe Bildungsberichterstattung 2008, 99), and there is need for a better coordination between the two data sources. While there is some information available on transitions to employment after the end of particular measures (“integration into the labor market”), there is no comprehensive and permanent data source available which records individual experiences in this system over time and links it to other forms of education and training over time.

The lack of transparency regarding empirical information about the “transition system” has further consequences: as young people in these measures do not appear as applicants in other sectors, it is very difficult to assess the overall demand in the VET system. For this part of VET it is especially important to know more about the incidence, the temporal extension, and the success of the measures. As long as there is no possibility of matching information across various years of data collection, information on previous experiences in the transition system needs to be collected when individuals enter training in the dual system or school-based training (or another training measure). In order to relate this information to the relevant populations of former participants in these measures, however, it would be most important to have information on the *timing* of these experiences. Again, it would be essential to collect information on the individual migration background.

#### *(4) Additional data sources*

Apart from data that have been designed primarily to inform about VET, there is a number of other (representative) data sources which do not have such a specific focus, but nevertheless carry relevant information. Most prominent example is the *Microcensus* which is a valuable source regarding the distribution of qualifications, differentiated by individual characteristics. In contrast to many other data sources, recent waves carry rather comprehensive information about both nationality and migration background. Conceptualized as a short-term panel it is in principle able to map educational transitions. There are also a few direct indicators in the questionnaire, which could be extended. To give but one example: As of 2005, information on additional vocational degrees is available; however, this is restricted to (higher education) graduates, so it does not allow identifying multiple training episodes in secondary vocational training. With regard to such additional data sources, refining and harmonizing the definitions of specific variables may greatly enhance their compatibility and hence their value for empirical research on training issues. In general, conceptualizing and further developing these data sources should be done in close cooperation with potential users; a good way would be issue-specific expert workshops.

### **3. Questions of data access and use**

Collecting adequate data is crucial, but for research in practice, *data availability and access* are other central issues. In recent years, promising progress has been made regarding the (projected) systematic and regular collection of process-based adequate, individual-level data

on VET. However, up to this point is not clear to what extent and in which form this stock of information will be accessible to scientific researchers. There is great need for a well-regulated access to these data (including their systematic documentation). Major advances can be expected from the recently set up *Forschungsdatenzentrum* for vocational training data at the Bundesinstitut für Berufsbildung (BIBB-FDZ). Relevant data on VET are held by very different institutions, so a common and regularly updated directory would be helpful for researchers. Another aspect concerns the range of available databases beyond official statistics. While data access and support for established surveys like the GSOEP are exemplary, many issue-specific datasets never become visible or available to potential users because the principal investigators are obliged by their funding contract to delete the data once the project is finished. The rule for projects commissioned by public authorities should be that the collected data should be made available through the relevant data archive or *Forschungsdatenzentrum*.

#### **4. Summary and recommendations**

Contemporary, individualized knowledge societies require comprehensive, up-to-date, and dynamic data, i.e. regularly collected data on individual histories of education and training rather than aggregate snapshots. Against this background, the current data situation concerning VET can be summarized as follows:

Information about the *Dual System* of apprenticeship training is comparatively comprehensive, but the positive change towards an individual-level accounting system in the statistics should be complemented by a systematic decision which allows studying longer sequences in training careers rather than single transitions. The simplest solution for this would be a permanent ID number. The information basis regarding *full-time vocational schooling* is much smaller. This situation is unsatisfactory, not least against the background of the gender-specific participation in these institutions. It is also rather heterogeneous due to the federalist structure of the German school system, and further efforts of coordination are necessary to build up a regular accounting system which allows dynamic analyses on the individual level. There is still no systematic reporting concerning the “*transition system*” of measures of vocational qualification which can be related to transitions to regular forms of training.

In spite of the progress which has undoubtedly been made regarding these various data sources, a decisive deficit remains: the fact that the three major sectors of the VET system are

covered by very different regulations and procedures of data collection. This means that transitions between these sectors can be analyzed only in a very selective and limited way. Moreover, most interesting research questions regarding the links between the VET system and *other* educational institutions (i.e., the general school system or higher education) as they are represented in patterns of educational careers can still not be investigated on the basis of process-produced data, though it is known from survey research that careers in education and training do transcend the borders of specific institutions and levels. Linking data bases on different areas of education and training remains therefore one of the top points on the future agenda, and again, the introduction of a common ID number – in combination with a consistent concept of data protection and data availability – could greatly help to overcome this deficit. Moreover, great care should be taken to harmonize the definitions and rules of data collection in the various sectors of the education and training system.

Hence, there are major opportunities to improve the scope and the quality of regularly collected, process-based data. Nevertheless, in the foreseeable future also (large-scale) surveys will be indispensable and therefore require adequate funding. They are necessary to get data on specific issues and for studies interested in causal questions. Such analyses require a broader range of individual characteristics including other aspects of life and information about parents and families: Educational careers are embedded in social relations, and the impact of families and life situations on educational decisions is strong. Moreover, it is necessary to include all three sectors (i.e., also the “transition system”) into a systematic data collection of competencies, most prominently in the multi-cohort *National Educational Panel Study* (NEPS), which is about to start in 2009/2010. Given its sample size and scope as well as the comprehensiveness of its data collection program, NEPS can be expected to accomplish an unprecedented level of integration among analyses focusing on various institutional stages of educational and training careers.

Additional demands arise from the expected trend towards international comparisons in all areas of education and training. These comprise large-scale (cross-sectional) reporting, longitudinal studies and evaluation studies of specific institutions or measures. VET has become part of such endeavors at a relatively late stage, but large-scale assessments in analogy to PISA (VET-LSA: Baethge et al. 2006; Programme for the International Assessment of Adult Competencies (PIAAC)) have been projected. Coordination among various programs assessing both general education and vocational and academic training may provide new challenges. Cross-national research has shown that it is often more salient to compare whole *systems* of education and training rather than specific elements defined on the

basis of nominal institutional classifications. To allow such a “system evaluation”, it is again important to link data from different parts of the education and training system and to harmonize the definitions used when collecting these data – on both the national and the international level.

## References:

- Autorengruppe Bildungsberichterstattung (2008): Bildung in Deutschland 2008. Ein indikatorengestützter Bericht mit einer Analyse zu Übergängen im Anschluss an den Sekundarbereich I. Bielefeld.
- Baethge, M./Buss, K.-P. and Lanfer, C. (2003): Konzeptionelle Grundlagen für einen Nationalen Bildungsbericht: Berufliche Bildung und Weiterbildung/Lebenslanges Lernen. Bonn/Berlin.
- Baethge, M./Buss, K.-P. and Lanfer, C. (Eds.) (2005): Expertisen zu den konzeptionellen Grundlagen für einen Nationalen Bildungsbericht. Bonn/Berlin.
- Baethge, M./Achtenhagen, F./Arends, L./Babic, E./Baethge-Kinsky, V. and Weber, S. (2006): Berufsbildungs-PISA: Machbarkeitsstudie. Stuttgart.
- Bellmann, L. (2005): Der Stand der Aus- und Weiterbildungsstatistik in Deutschland. In: Baethge et al., 67-101.
- Brosi, W. (2005): Anmerkungen zur Verfügbarkeit von Statistiken im Bereich der beruflichen Bildung als Basis für eine umfassende Bildungsberichterstattung. In: Baethge et al., 103-140.
- Bundesministerium für Bildung und Forschung (BMBF) (2007): Berufsbildungsbericht 2007. Berlin.
- Dietrich, H./Behle, H./Böhm, R./Eigenhüller, L. and Rothe, Th. (2002): Design der IAB-Begleitforschung zum Jugendsofortprogramm im Kontext des IAB-Forschungsschwerpunkts School-to-Work-Transition. Nürnberg.
- Hillmert, S. and Jacob, M. (2003): Bildungsprozesse zwischen Diskontinuität und Karriere: das Phänomen der Mehrfachausbildungen. Zeitschrift für Soziologie 32, 4/2003, 325-345.
- Krüger, H. (2005): Zur Datenlage vollzeitschulischer Berufsausbildung. In: Baethge et al., 141-164.
- Statistisches Bundesamt (2007a): Berufsbildungsstatistik – Begriffe und Erläuterungen. Wiesbaden.
- Statistisches Bundesamt (2007b): Fachserie 11, Reihe 2: Bildung und Kultur – Berufliche Schulen. Wiesbaden.
- Troltsch, K./László, A./Bardeleben, R. von and Ulrich, J.G. (1999): Jugendliche ohne Berufsausbildung. Eine BiBB/EMNID-Untersuchung. Bonn.
- Uhly, A. (2006): Weitreichende Verbesserungen der Berufsbildungsstatistik ab April 2007. In: Krekel, E.M./Uhly, A. and Ulrich, J.G. (Eds.): Forschung im Spannungsfeld konkurrierender Interessen. Die Ausbildungsstatistik und ihr Beitrag für Praxis, Politik und Wissenschaft. Bonn.
- Weishaupt, H. and Fickermann, D. (2000): Informationelle Infrastruktur im Bereich Bildung und Kultur, Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. Erfurt.