Economics Letters 141 (2016) 73–76

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

# How much can we trust maternal ratings of early child development in disadvantaged samples?

Malte Sandner [a,*], Tanja Jungmann [b]

[a] NIW Hannover, Königstrasse 53, 30175 Hannover, Germany
[b] Universität Rostock August-Bebel-Strasse 28, 18051 Rostock, Germany

## HIGHLIGHTS

- We analyze the validity of a screening questionnaire (SQ) to assess child development.
- The present study uses samples of disadvantaged mothers for validation.
- We find significant correlations between the SQ ratings and those of professional examiners.
- However, concordance is smaller for strongly disadvantaged mothers.

## ARTICLE INFO

*Article history:*
Received 16 September 2015
Received in revised form
14 January 2016
Accepted 14 January 2016
Available online 15 February 2016

*JEL classification:*
J13
C42
C91

*Keywords:*
Child development
Validation of survey measures
Disadvantaged mothers

## ABSTRACT

An increasing number of panel studies use short screening questionnaires to assess infant development. Although some research examines the validity of screening questionnaires for middle-class families, knowledge about their accuracy in disadvantaged households is scarce. This paper validates a short screening questionnaire included in the German Socio-Economic Panel (SOEP) with the Bayley Scales of Infant Development (BSID) as an external criterion with a disadvantaged population. The results reveal significant correlations between the screening questionnaire ratings and the BSID scores for disadvantaged mothers. However, the concordance of maternal ratings and test results decreased in mothers with multiple risk burdens.



## 1. Introduction

Recent evidence from economic research underlines the crucial role of early childhood in lifelong human capital accumulation (e.g., Cunha and Heckman, 2007; Almond and Currie, 2011). Many economists claim that policies for disadvantaged families starting in early childhood are efficient to reduce poverty and inequality (e.g., Heckman and Masterov, 2007). To understand the mechanisms of the effects of these policies, it is important to consider both, long-term outcomes, such as income or grades, and short-term outcomes, such as infant development. Field experiments offer the opportunity to assess infant development using tests administered by professional examiners, with the Bayley Scales of Infant Development (BSID; Bayley, 1993) considered the "gold standard". However, developmental tests are costly to conduct and difficult to administer in large panel studies. Therefore, panel studies such as the German Socio-Economic Panel Study (SOEP) or the National Longitudinal Survey of Children and Youth (NLSCY) use less time and cost intensive screening questionnaires (SQs) in which the caregiver rates the infants development.

However, measuring children's skills by asking the parents may be accompanied with many threats to validity. Caregivers can be positively or negatively biased in their perception of their children, may give socially desired answers, or may report developmental achievements only because they are asked for (e.g., Schwarz, 1999). Furthermore, child development is a highly dynamic process with large individual variation. While some evidence indicates that these threats only slightly bias the screening scores in average populations (e.g., Johnson et al., 2004), research has not determined whether this finding also applies to disadvantaged mothers, who are the main target population for early childhood interventions. The described threats may be more severe for disadvantaged mothers because they are often less knowledgeable about the milestones of child development. Additionally, they are more often socially isolated and thus have limited

* Corresponding author.
*E-mail address:* sandner@niw.de (M. Sandner).

http://dx.doi.org/10.1016/j.econlet.2016.01.021
0165-1765/

occasions to compare their own children's development to that of age-equivalent children. Even if they have a functioning social network, these mothers are often too burdened with daily crises and existential problems to fully realize developmental delays in various domains of child development.

This paper examines the validity of the SOEP-SQ (see Schmiade et al., 2008, for detailed description) in a sample of psychosocially disadvantaged mothers. The mothers exhibit between one and 12 social risk factors (e.g., low education, teenage pregnancy, isolation, experiences of violence or health problems). For validation, infants were screened using the SOEP-SQ and were tested using the BSID-II (German version by Reuner et al., 2007) as an external criterion. The results revealed significant correlations between the overall disadvantaged mothers' SOEP-SQ ratings and the BSID-II scores ($r = 0.35$–$0.56$). We found that the concordance between the maternal ratings and the test results is lower for mothers with five or more risk factors than for mothers with less than five risk factors, although our sample size lacks power to show that these differences are significant. Overall, the results indicate that SQs draw a valid and reliable picture of child development, also in a disadvantaged sample. However, studies evaluating childhood policies for families with multiple risk factors should be cautious when using SQ results.

## 2. Measuring child development

Psychologists distinguish between two procedures to measure child development: screening and tests. Following the definition of the American Academy of Child and Adolescent Psychiatrists, screening involves the identification of current behavioral health or developmental status. In the screening procedure, developmental achievement in various domains (mental, psychomotor, behavior, and language) is typically documented using a standardized questionnaire. The family member or caregiver is expected to be a reliable source who knows the child well. Typically, the screening is quick and easy to administer and score. In contrast, testing is a more comprehensive process, as it involves the family and evaluates the strengths and difficulties of the child and the family in all life domains. A qualified individual with the appropriate credentials required by the licensing authority conducts the assessment of developmental status.

The SOEP-SQ is a short version of the Vineland Adaptive Behavior Scale (VABS; Sparrow et al., 2005), which is a screening procedure that obtains a parent's report of the child's adaptive behavior from one month of age through adolescence. Parents are asked to indicate whether their child always (2), sometimes (1), or never (0) demonstrates age-normed adaptive behaviors in four critical domains of functioning (communication, daily living skills, motor skills, and socialization). The interview contains 297 items and requires 30–60 min. The SOEP uses a modified short version of the VABS with the four subscales of "Speaking", "Everyday Skills", "Movement", and "Social Relationships" for infants between 24 and 36 months of age. This version includes a total of 20 items that have not been validated using external criteria.[1]

In contrast to the screening approach, the BSID-II is an examiner-administered test of infant development that assesses the mental abilities and psychomotor abilities of infants between the ages of one and 42 months. Mental abilities include perceptual acuities, acquisition of object constancy, memory, learning and problem solving, nonverbal and verbal communication, complex language, and abstract thinking. Psychomotor abilities are defined as the degree of body control, large muscle coordination, fine motor skills of the hands and fingers, dynamic movements, postural imitation, and the ability to recognize objects by the sense of touch with different items. The BSID-II comprises one scale to assess the mental abilities (Mental Development Index, MDI) and another scale to evaluate the psychomotor abilities (Psychomotor Development Index, PDI).

In conducting the BSID-II, the examiner presents a series of play materials or pictures to the child. She provides brief instruction for each task and observes and documents whether the child is able to solve it. The test sessions are videotaped and scored after the interview by a developmental psychologist. Raw scores for each subtest are summed and transformed into standard scores (MDI and PDI) with a mean of 100 and SD of 15. Because of the comprehensive and appropriate content, construct, and criterion validity, the BSID-II represents the "gold standard" of neurodevelopmental assessment in infancy and toddlerhood (e.g., Tylenda et al., 2007). Regarding predictive validity, most evidence refers to the MDI. In their literature review, Fryer and Levitt (2013) show that the average correlation between MDI and future IQ is approximately 0.3 at 9 years of age. For Germany in particular, Coneus et al. (2012) find a positive correlation between MDI and high school graduation. Less is known about the predictive validity of PDI.

## 3. Data

We used data from the evaluation of the *Pro Kind* home visiting program to determine the validity of the SOEP-SQ. *Pro Kind* is an early childhood intervention in which 755 psychosocially disadvantaged first-time mothers and their families received home visits from pregnancy until the child's second birthday. Participating mothers exhibited between one and 12 social risk factors and lived in a household that received social welfare benefits or unemployment compensation, had an income as low as social welfare benefits, and/or was over-indebted. Only mothers between the 12th and 28th weeks of pregnancy were affiliated with the project between November 2006 and December 2009. *Pro Kind* is located in three German federal states (two in West and one in East Germany). Jungmann et al. (2009) and Sandner (2013) provide detailed information about the program and the characteristics of the sample.

An interdisciplinary research team administered the BSID-II at six, 12 and 24 months after the child was born. Additionally, biannual telephone interviews surveyed maternal income, employment, fertility and health. The telephone interview conducted shortly after the second birthday of the child contained the SOEP-SQ. The BSID-II and SOEP-SQ were conducted at approximately the same time but in two different interviews, thus minimizing problems of induced responses.

Table 1 shows the descriptive statistics of the sample that we used to validate the SOEP-SQ.[2] The first rows present the outcomes for the SOEP-SQ and BSID-II. *Pro Kind* children scored lower than comparable children on the SOEP-SQ (Schmiade et al., 2008), and below the norm of 100 on the MDI and PDI. The last rows in Table 1 present the demographic factors of the children and the mothers. The children were approximately the same average age when they were tested with the BSID-II and when their mothers answered the SOEP-SQ. At their first child's birth, the mothers were young, had a low socioeconomic status indicated by a mean of 5.4 risk factors and a low educational background with 60% of the mothers in the lowest education group. These characteristics are potential obstacles to child development, as reflected in the children's low scores on the BSID-II scales and the SOEP-SQ maternal ratings.

[1] Appendix provides the full text (English translation) of these SOEP screening questions.

[2] Children were included in the sample if their mothers answered the SOEP-SQ when their child's age ranged from 23 to 27 months, resulting in a sample of 166 children.