Early Childhood Research Quarterly 48 (2019) 173–185

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

# Effects of home-based early intervention on child outcomes: A randomized controlled trial of Parents as Teachers in Switzerland

Simone Schaub[a,*], Erich Ramseier[b], Alex Neuhauser[a], Susan C.A. Burkhardt[a], Andrea Lanfranchi[a]

[a] University of Applied Sciences in Special Needs Education, Zurich, Switzerland
[b] PHBern—University of Teacher Education, Berne, Switzerland

ARTICLE INFO

*Article history:*
Received 18 June 2018
Received in revised form 6 March 2019
Accepted 25 March 2019
Available online 6 May 2019

*Keywords:*
At-risk families
Early childhood development
Early intervention
Home-visiting program
Parents as Teachers
Randomized controlled trial

ABSTRACT

This randomized controlled trial examined the effectiveness of the Parents as Teachers (PAT) program in the German-speaking part of Switzerland. During the first three years after birth, 132 at-risk families with new-born children were supported by PAT with regular home visits and group connections. The 116 families in the control group had access to the normal community services, but were not supported by PAT. Baseline data and annual outcome data on health, adaptive behavior, developmental status, cognitive skills, language skills, motor skills, problem behavior, and effortful control were collected. As shown by Generalized Estimating Equations, PAT improved children's adaptive behavior, developmental status, and language skills at the age of 3 years. Problem behavior was reduced in families with the highest risk. The results are discussed in the light of Bronfenbrenner's bioecological model.

© 2019 Elsevier Inc. All rights reserved.

## 1. Introduction

Early childhood home-visiting programs support at-risk families (e.g., affected by psychosocial risk factors such as social isolation, poverty, or mental illness) during the first years of their children's lives, aimed at remediating the negative impact of such risk factors on child development (e.g., Rutter, 2011; Sameroff, 2010). The adverse impact of psychosocial risk factors has been demonstrated repeatedly regarding different aspects of child development and wellbeing (e.g., Duncan & Brooks-Gunn, 1997; Hertzman & Boyce, 2010; Rutter, 2000). Large-scale longitudinal studies that followed new-born children for up to 40 years have demonstrated lifelong adverse effects on physical, academical and social outcomes (Laucht, Esser, & Schmidt, 1997; Schoon, 2006; Sroufe, Egeland, Carlson, & Collins, 2005; Werner & Smith, 2001). At the same time, these studies emphasize protective factors and their interplay with risk factors, which in turn foster children's resilience. Home-visiting programs draw on these findings and seek to reinforce protective factors such as social support or family values and goals (Ramey & Landesman Ramey, 1993).

Meta-analytic evidence is mixed regarding the effectiveness of home-visiting programs on child outcomes (Rayce, Rasmussen, Klest, Patras, & Pontoppidan, 2017; Sweet & Appelbaum, 2004; Taubner, Wolter, & Rabung, 2015). Effect sizes range from −.49 to .69, with a mean of .17 of a standard deviation (Karoly, Kilburn, & Cannon, 2005). However, especially high-quality programs, such as the Nurse Family Partnership (NFP), have shown positive effects on child cognitive, language and behavioral outcomes (Olds, Sadler, & Kitzman, 2007). Outside the United States, there is little evidence of the effectiveness of home-visiting programs (Rayce et al., 2017; Taubner et al., 2015). Hence, findings mostly stem from studies conducted in the U.S., where the first support programs for disadvantaged children and their families were implemented in the 1960s (e.g., Ramey, Sparling, & Landesman Ramey, 2012; Schweinhart et al., 2005).

Switzerland has a high standard of universal care for families with newborns. It includes home visits by midwives during the immediate postnatal period, parental and educational counselling, and regular medical check-ups. Likewise, early special education is well established for children with disabilities or developmental delays. In contrast, the secondary preventive measures relevant here are undersupplied, despite there being an undeniable need for them (Stamm et al., 2009). Children's well-being and educational opportunities are strongly dependent on family background (Burger & Walk, 2016; OECD, 2016). In response, there have been

* Corresponding author at: University of Applied Sciences in Special Needs Education, Schaffhauserstrasse 239, P.O. Box 5850, 8050 Zurich, Switzerland.
*E-mail address:* simone.schaub@hfh.ch (S. Schaub).

https://doi.org/10.1016/j.ecresq.2019.03.007
0885-2006/© 2019 Elsevier Inc. All rights reserved.

individual local initiatives aimed at at-risk families in recent years, including home-visiting programs (Diez Grieser & Simoni, 2012). The insufficient care of disadvantaged children of pre-school age mainly holds for the German-speaking part of Switzerland and not the French- and Italian-speaking regions (Burger, 2013). Congruously, the unequal distribution of educational opportunities proves to be common issue in all German-speaking European countries (OECD, 2016). Three characteristics in the education systems of these countries are of particular importance in terms of equity (Crul & Vermeulen, 2006): the stratified school system, a late start of compulsory schooling, and the comparatively high expense of preschool education. Germany established the "National Center for Early Prevention" (NZFH, www.nzfh.de) as part of a federal initiative in 2007, which funds the implementation of secondary preventive measures and their efficacy research (Cierpka & Evers, 2015). To date, these studies have shown unsatisfactory efficacy with regard to home-visiting programs (see Taubner et al., 2015, for a meta-analysis). In particular, the one study using an experimentally controlled research design found treatment effects of the NFP program on child outcomes in only one subgroup, namely the most vulnerable families (Jungmann et al., 2015; Sierau et al., 2016). Other European countries, such as Great Britain, have similarly reported non-effective implementations of the NFP program (Robling et al., 2016). However, before a widespread introduction of a U.S. program it is essential to have scientific evidence showing that the program is more effective than the existing support services (Olds, 2016).

### 1.1. Parents as Teachers

PAT is a parent-training program beginning during pregnancy or shortly after birth, and lasting until the child's third birthday, which was founded in 1984 in the State of Missouri (https://parentsasteachers.org/about/). U.S. trials have demonstrated the program's effectiveness in improving child outcomes (see Neuhauser, 2014, for a review), and the program meets the Home Visiting Evidence of Effectiveness (HomVee) standards (Sama-Miller et al., 2017). However, failure to replicate the positive outcomes in more recent trials highlights the necessity to replicate the effectiveness in rigorous randomized controlled trials (RCT; Olds et al., 2007). PAT was first implemented in a European context in Nuremberg in Germany as part of a municipal integration program in 2004, and subsequently extended to other cities in Germany and Switzerland (www.pat-mitelternlernen.org). However, the effectiveness of the program in Germany or Switzerland has to date not been investigated.

The PAT program encompasses the following goals: (a) the increase of parental knowledge of early childhood development and the improvement of parental practices, (b) the early detection of developmental delays and health issues, (c) the prevention of child abuse and neglect, (d) the long-term increase in children's school readiness and success. Four program components frame the means by which these goals are to be achieved.

(1) The core component of the program is home visits, which are carried out by qualified parent educators with a higher degree in early education. The educators in the current project were pediatric nurses with additional training as parent counselors and one midwife. All had completed the weekly basic training as parent educators in Nuremberg, as well as attending further trainings and the annual recertifications. Home visits should ideally take place in the first three years—or at least the first two years—with a minimum of 10 visits per year. In the current project 89.3% of the families participated at least 24 months, and 95.9% were visited 10 times or more per year. Each home visit requires three areas to be addressed on the basis of a curriculum: development-oriented parenting, parent-child interactions, and the well-being of the family.
(2) Group connections take place once a month. PAT does not specify the minimum number of meetings to be attended. In the current study 52.7% of the families attended at least four group connections per year. The other half participated less often, or not at all (5.5%). These connections serve to promote the networking of parents and the provision of information on educational practices, parent-child interactions, and community services for families.
(3) Screenings on general health development and on hearing and vision development take place at least once a year.
(4) The last component is the support of the parents in networking in the community and the referral to other public institutions and community services, as needed.

The curriculum was translated and adapted to German conditions in 2004. Special consideration was given to German-as-a-second language families, including providing parent information in multiple languages. Also, in the current project home visits and group connections were carried out with the assistance of intercultural interpreters in cases where German-language skills proved insufficient. The umbrella organization in Nuremberg formulates model fidelity requirements that affiliates must meet. In particular, providers must produce annual performance reports on process quality and program impact.

### 1.2. Theoretical and empirical background

Parents as Teachers models the understanding that the early environment is crucial for healthy child development (e.g., Sameroff, 2010; Shonkoff, 2010) and is based on Bronfenbrenner's bioecological model (Bronfenbrenner, 1986). This model posits that the developing child forms a dynamic unit with the environment in which each influences the other. Continuous, reciprocal, and increasingly complex interactions within this unit, so- called proximal processes, are the "primary engines" of development (Bronfenbrenner & Morris, 2006, p. 798). Characteristics of the developing person, the developmental outcome, the environment, and the time period influence these processes. At the core of the model is the child and its biological and psychological resources, with which it shapes its own development. Developmentally generative characteristics such as curiosity or engagement initiate and maintain proximal processes; developmentally disruptive characteristics such as a lack of behavioral control or—conversely—a prevalence of apathy hinders the occurrence of proximal processes. These characteristics are not only dispositions of the child, but also stand for two opposing developmental outcomes, whereby developmental competence or the "acquisition and further development of knowledge and skills" is opposed to developmental dysfunction or "the recurrent manifestation of difficulties on the part of the developing person in maintaining control and integration of behavior" (Bronfenbrenner & Morris, 2006, p. 803). Physical and psychosocial qualities of the immediate environment (i.e., microsystem) such as the availability of learning materials or parenting behaviors, the associated network of the family (mesosystem), distal qualities of the exosystem such as health care provision, and, finally, the cultural (macrosystem) and historical and contemporary context (chronosystem) can have direct or indirect influences on development. In the early years, proximal processes mainly occur in the microsystem of the home environment and family–child interactions. Home-visiting programs aim to improve this environment and parenting practices in order to indirectly influence proximal processes and ultimately enhance child development (Peterson, Luze, Eshbaugh, Hyun-joo, & Kantz, 2007).

A risk factor is "any individual or environmental factor associated with the increased likelihood of developing negative or undesirable outcomes" (Evans, Li, & Whipple, 2013, p. 1342). Developmental outcomes differ to a varying degree as a function of biological and psychosocial risk factors (Shonkoff & Phillips, 2000). The Mannheim Study of Children at Risk in Germany found that motor development was primarily predicted by biological risk (i.e., prenatal and perinatal complications), while psychosocial risk (e.g., low educational levels) substantially predicted problem behavior and cognitive development (Coneus, Laucht, & Reuss, 2012; Laucht et al., 1997). In terms of developmental competence, because it heavily relies on the level of daily exposure to spoken language and cognitive stimulation, language development is particularly susceptible to experience (Clark, 2016). Poor exposure to language results in lower vocabulary and language growth (Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010). Experience, parental stimulation, and parent-child interactions play an equally important role in the development of early cognitive skills (Ayoub et al., 2009; Tucker-Drob & Harden, 2012), whilst motor development, by contrast, is mainly dependent on maturation and requires little parent–child interaction (DiPietro, 2000). As with competence, developmental dysfunction is malleable through experience and parent–child interactions: negative parental control may lead to impoverished self-regulatory skills (Karreman, van Tuijl, van Aken, & Dekovic, 2006) and problem behavior (Olson, Ceballo, & Park, 2002). According to the bioecological model,

> the greater developmental impact of proximal processes on children growing up in disadvantaged or disorganized environments is to be expected to occur mainly for outcomes reflecting developmental *dysfunction*. By contrast, for outcomes indicating developmental *competence*, proximal processes are posited as likely to have greater impact in more advantaged and stable environments. (Bronfenbrenner & Morris, p. 803, italics in original)

This assumption is based on the following reasoning. Stable and organized environments are needed to enable parents to engage with developmental competencies in a supportive and responsive manner. The absence of such an environment results in a lack of resources—be it psychological, physical, or educational—to provide stimulating experiences. Research evidence, such as the finding that at-risk families offer dramatically less verbal input than the average family (Hart & Risley, 1995), supports this claim. In addition to insufficient parental resources, children from at-risk families show more manifestations of developmental dysfunction themselves than children from privileged families (Olson et al., 2002). This combination results in disadvantaged parents focusing more strongly than privileged parents on childhood dysfunction as opposed to competence. To the best of our knowledge, however, this assumption awaits empirical evidence. On the one hand, differential effects of proximal processes on developmental competence versus dysfunction over the course of the early years have yet to be investigated simultaneously. On the other hand, previous intervention studies show no evidence of greater impact on developmental dysfunction as opposed to competence in disadvantaged families (see Gomby, 2005, for a review).

A pivotal final consideration is the question what it is that makes the environment an at-risk setting. The bioecological model assumes that multiple risk factors have a compound escalating effect since similar disruptive properties of interconnected microsystems tend to reinforce each other (Bronfenbrenner & Morris, 2006). Robust empirical evidence supports this claim. The presence of a single risk factor rarely affects child development. Only the exposure to multiple risk factors increases the likelihood of developmental consequences (see Evans et al., 2013, for a review). These findings are central to the implementation of home-visiting programs because they apply to their primary addressees. However, the current state of research does not yet provide sufficient evidence to answer the question for which families home-visiting programs are best suited, leading Gomby, Culross, and Behrman (1999) to urge researchers to further explore which families benefit most from these programs. NFP program has proven to be effective for a very specific subgroup, namely young, first-time mothers (Olds et al., 2007). As a universal preventive measure for a larger population, there is so far a lack of findings on PAT concerning the adequate target group. Adding to this, the notion of risk raises a further fundamental concern. Children in Switzerland are less disadvantaged in terms of poverty or access to healthcare than children in the U.S. (OECD, 2018). At the level of the exosystem, the environments in which the children grow up in can therefore not be directly compared. Previous replications in Western European countries found either no positive impact on child development (Robling et al., 2016, for the United Kingdom) or only in a subgroup (Sierau et al., 2016, for Germany). Therefore, robust evidence of the effectiveness of home-visiting programs in Switzerland or other Western European countries is highly desirable.

In sum, two important research gaps were identified. First and foremost, disadvantaged children in Switzerland suffer measurable inequalities, but the effectiveness of home-visiting programs as suitable means of promoting equity has not been substantiated to date. Secondly, studies to date do not provide sufficient information on which microsystems and for which developmental outcome home-visiting programs are suitable and effective.

### 1.3. The current study

Bronfenbrenner's process-person-context-time (PPCT) model (Bronfenbrenner & Morris, 2006) forms the theoretical framework of the current research project. The longitudinal RCT ZEPPELIN 0—3 was conducted to analyze the effectiveness of the PAT program in the German-speaking part of Switzerland. The study consisted of four measurement points over the course of the first three years. At each measurement point child and context characteristics, as well as child-context interactions, were assessed (Lanfranchi & Neuhauser, 2013). The purpose of the current paper is twofold: (1) to examine the effectiveness of PAT in the German-speaking part of Switzerland on child outcomes after three years of intervention, and (2) to identify families for which PAT is effective. Child health and adaptive behavior were monitored over the entire period. whereby the latter included age-appropriate developmental milestones and self-help skills. Effects on developmental competence were examined on the basis of three skill domains: cognitive, language, and motor. Problem behavior and effortful control were assessed as manifestations of developmental dysfunction. Based on the aforementioned theoretical and empirical evidence, it was hypothesized, that (1) participation in PAT improves language and cognitive skill outcomes and reduces problem behavior and lack of control, and that (2) PAT is most effective in high-risk families.

## 2. Methods

### 2.1. Recruitment and randomization

For recruitment purposes, an interdisciplinary network within the existing community-service infrastructure (e.g., parent-counseling offices, pediatricians, midwives, social counseling, psychological and psychiatric services) was established. This network recruited families exhibiting at least two distinct psychosocial risk factors on the personal (e.g., mental illness), the family (e.g., single parents), the social (e.g., no social network), or on the

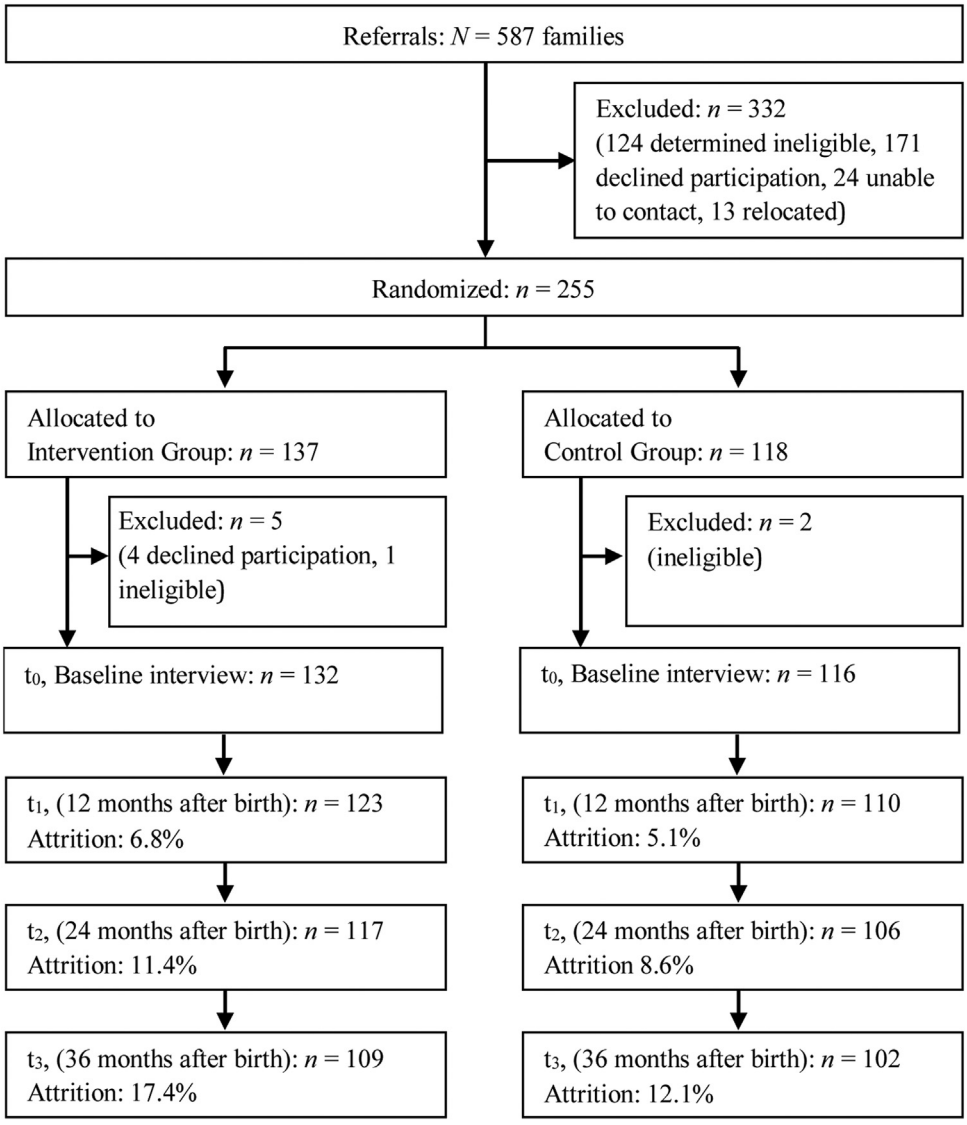


**Fig. 1.** CONSORT flow diagram.

material level (e.g., confined living space). Exclusion criteria were persons with (a) no permanent residency permit, (b) severe illness or disability of the child, (c) severe illness or disability of the parent requiring inpatient and long-term psychiatric treatment, and (d) other intensive treatments or child protection procedures. These families were then referred to the study on the voluntary basis ($N = 587$, see Fig. 1). In Switzerland, the parent-counseling offices receive notification of newborns in the area and contact all families by standard. If they were not able to contact a family, additional measures—usually the visiting of the family at home—were undertaken. A parent educator then contacted the selected families, visited them at home, and informed the parents about the control-group design and the study objective to evaluate the implementation of PAT as a prospective new support service. Of the total, a remaining 255 families gave their consent to participate and were referred to randomization. Stratified block randomization was used to assign each family to either the intervention group (IG) or the control group (CG; see Neuhauser, Ramseier, Schaub, Burkhardt, Templer, & Lanfranchi, 2015 for a detailed description). Strata were project site, cumulative psychosocial risk factors (high/low), marital status (single parent: yes/no), and German-language proficiency (interpreter: yes/no).

### 2.2. Participants

After randomization, four families declined participation and three families were later determined ineligible due to severe illness of their children, which only became evident with increasing age (e.g., one premature child with severe developmental delays). Therefore, the final experimental group consisted of 248 families ($N = 261$ children, including 13 twins). Table 1 shows the socio-demographics of the participating families. What is noteworthy is that the primary language of 78.3% of the children was non-German. The parents were from different ethnic backgrounds, the most frequent being Swiss nationals (mothers, 26.6%, fathers, 29.0%) Portuguese (8.9%, 9.7%), Turkish (8.5%, 9.3%), Kosovar (8.5%, 10.5%), and Eritrean (5.2%, 4.8%). The indicators of multi-ethnicity used were: (a) children's primary language, (b) children's bilingualism, (c) the mothers' German proficiency, and (d) the mothers' life share in Switzerland (ascertained by dividing the number of years lived in Switzerland at the time of the child's birth by the mothers' age, multiplied by 100). Baseline equivalence was analyzed using binary logistic regression on group assignment. This analysis showed no statistically significant effects of the socio-demographics listed in Table 1. However, bivariate comparisons revealed marginally sig-

**Table 1**
Demographic characteristics at baseline assessment and at 36 months: comparison of the experimental groups and risk levels.

| | Baseline | | | | | 36 months | |
|---|---|---|---|---|---|---|---|
| | IG (=1) | CG (=0) | Low-risk | Medium-risk | High-risk | IG | CG |
| *n* | 132 | 116 | 64 | 111 | 73 | 109 | 102 |
| Variable | %/*M* (*SD*) | %/*M* (*SD*) | %/*M* (*SD*) | %/*M* (*SD*) | %/*M* (*SD*) | %/*M* (*SD*) | %/*M* (*SD*) |
| Children | | | | | | | |
| Girls (=1) | 56.8 | 48.4 | 47.8 | 52.6 | 57.7 | 55.8 | 50.0 |
| Birth weight (gram) | 3171.9 (611.9) | 3289.0 (593.4) | 3330.9 (632.7) | 3204.2 (541.8) | 3169.4 (665.3) | 3214.6 (606.0) | 3302.5 (612.2) |
| Gestational age (weeks) | 39.0 (2.2) | 39.3 (1.9) | 39.1 (2.1) | 39.3 (1.6) | 38.9 (2.6) | 39.0 (2.2) | 39.3 (2.0) |
| Age at randomization (days) | 51.3 (38.9) | 55.0 (49.8) | 54.9 (42.9) | 58.5 (40.2) | 43.1 (50.0)$^{+}$ | 53.3 (35.9) | 56.5 (50.7) |
| Foreign first language (=1) | 377.3 | 79.3 | 73.1 | 79.3 | 79.5 | 77.1 | 81.4 |
| Bilingual (=1) | 14.4 | 6.9$^{+}$ | 16.4 | 9.5 | 9.0 | 16.5 | 6.9$^{*}$ |
| Family | | | | | | | |
| Recruited through additional measures (=1) | 14.4 | 24.1$^{+}$ | 21.9 | 21.6 | 12.3 | 10.1 | 23.5$^{**}$ |
| Twins (=1) | 10.1 | 9.8 | 4.7 | 4.5 | 6.8 | 3.7 | 5.9 |
| Number of siblings | .6 (0.8) | .7 (0.9) | .4 (0.6) | .6 (0.9) | .9 (1.0)$^{**}$ | .6 (0.9) | .7 (1.0) |
| Global risk HBS | 47.5 (16.6) | 44.8 (15.6) | 25.7 (6.4) | 45.4 (5.5) | 65.7 (7.1)$^{***}$ | 47.1 (17.1) | 44.4 (15.7) |
| ISEI | 27.1 (21.2) | 31.5 (23.2) | 39.8 (24.3) | 29.7 (21.6) | 19.1 (16.2)$^{***}$ | 27.8 (22.2) | 32.5 (23.7) |
| Mother | | | | | | | |
| Age at birth (years) | 29.4 (5.9) | 29.9 (5.5) | 30.5 (5.3) | 29.8 (5.8) | 28.7 (5.9) | 30.1 (5.8) | 30.2 (5.6) |
| Single parent (=1) | 13.0 | 14.7 | 1.6 | 9.1 | 31.5$^{***}$ | 11.0 | 13.7 |
| No post-compulsory education (=1) | 31.8 | 38.2 | 20.3 | 41.8 | 54.8$^{***}$ | 39.4 | 44.6 |
| Life share in Switzerland (%) | 37.2 (36.5) | 40.9 (35.8) | 49.5 (40.6) | 34.6 (33.0) | 36.1 (35.1)$^{*}$ | 38.7 (36.6) | 39.4 (34.8) |
| German language proficiency | | | | | | | |
| None (=1) | 12.1 | 11.2 | 4.7 | 12.6 | 16.4 | 8.3 | 10.8 |
| Low (=2) | 19.7 | 18.1 | 17.2 | 21.6 | 16.4 | 20.2 | 19.6 |
| Sufficient (=3) | 28.0 | 22.4 | 20.3 | 26.1 | 28.8 | 29.4 | 24.5 |
| Good (=4) | 40.2 | 48.3 | 57.8 | 39.6 | 38.4 | 42.2 | 45.1 |

*Note.* ISEI = International Socio-Economic Index of Occupational Status (Ganzeboom, De Graaf, & Treiman, 1992). *N* was 261 in the case of gender and birth weight (individual children).
Coding of the corresponding categories in parentheses.
$^{***}$ $p < .01$.
$^{**}$ $p < .01$.
$^{*}$ $p < .05$.
$^{+}$ $p < .10$, two-tailed.

nificant higher proportions of monolingual children and families recruited through additional measures in the CG. These differences were accounted for in the subsequent analyses.

### 2.3. Intervention

The intervention was carried out at three sites in the suburbs of Zurich, Switzerland ($n = 100$, $n = 81$, and $n = 67$ families) by 11 parent educators, each accompanying 18–28 families, with a balanced distribution of IG and CG. PAT was not offered to the CG. However, they had access to standard health services and were referred to child-related institutions in the community if necessary, e.g., if the wellbeing of the child was uncertain or if the researchers suspected developmental delay.

### 2.4. Data collection and measures

Data were collected at four measurement points. Baseline data collection consisted of a home visit and a guided interview by a research assistant prior to the intervention, and took place approximately three months after birth. The subsequent measurement points were set at the approximate birthdays of the children: $t_1$ at 12 months, $t_2$ at 24 months, and $t_3$ at 36 months. During the first appointment, research assistants performed a guided interview at the families' homes, supplemented by parent questionnaires. At $t_3$, research assistants additionally conducted small effortful control experiments with the children. Research assistants were blind to the experimental condition at the baseline interview; however, they were not masked anymore to the experimental condition at $t_1$–$t_3$. During a second appointment at public family centers, pediatricians who were blinded to the experimental condition tested the children's development and questioned the parents about the child's health and adaptive behavior. A third appointment was added at $t_3$, in which trained and blinded student collaborators tested the children's intelligence, again at public family centers. Intercultural interpreters translated interviews and tests with parents who were not proficient in German, and all questionnaires were available in multiple languages.

#### 2.4.1. Family characteristics

Socio-demographic information was gathered during the interviews. The Heidelberger Belastungsskala (Heidelberg Stress Scale, HBS; Sidor, Eickhorst, Stasch, & Cierpka, 2012) was used to assess the family's level of psychosocial risk at the baseline interview. This assessment is based on a semi-structured interview concerning risk and protective factors (e.g., employment situation, social contacts) and on observations (e.g., apartment size, parent–child interaction). These findings subsequently lead to an assessment of the psychosocial risk level on a continuous scale ranging from 0 to 100 in four domains: material (e.g., confined living space), social (e.g., no family support), family/personal (e.g., single parenting), and child (e.g., prematurity). Finally, all risk and protective factors result in a global risk score. Concurrent, predictive, and discriminant validity were evaluated in a sample of 284 German at-risk families (Sidor et al., 2012). In the current study, three coders performed the assessments. To determine the interrater reliability, 30 families were randomly selected and assessed by all coders. Interrater reliability was moderate to good, with intraclass coefficients ranging from .64 to .84, depending on the domain. The scores from the primary coder were used for the analyses.

Based on the stratification block risk-level, three groups were formed for additional analyses: Families categorized as "low-risk" had a global score below 40 (i.e., risk cannot be fully compensated by existing protective factors, IG, $n = 29$, CG, $n = 35$); families with a score between 40 and 59 (i.e., considerably more burdened than unaffected domains, IG, $n = 60$, CG, $n = 51$) were categorized

as "medium-risk"; and families with a value of 60 and above (i.e., possibilities of undisturbed family functioning are rare, IG, $n = 43$, CG, $n = 30$) as "high-risk". Increasing risk was associated with more siblings, lower ISEI, a larger proportion of single mothers and of uneducated mothers, and migration of the mothers to Switzerland at an older age (see Table 1).

#### 2.4.2. Child outcomes

At $t_1$–$t_3$, the pediatric testers used a non-standardized checklist to assess children's *health* (e.g., hearing difficulties) and children's *adaptive behavior*, or "skills that have been learned by people to function in their everyday lives" (Tassé, 2009, p. 114). Adaptive behavior comprised age-specific *self-help skills*, such as children's sleeping or toileting behavior, and *developmental milestones* in motor (e.g., standing on one leg), language (e.g., word combination), communication (e.g., question formulation), and conceptual development (e.g., pretense play).

*2.4.2.1. Developmental competence.* At $t_1$–$t_3$ the Bayley Scales of Infant and Toddler Development was employed (BSID III; Bayley, 2006; Reuner & Rosenkranz, 2014). The BSID III is the most widely used test in infant developmental assessments and intervention research (Johnson, Moore, & Marlow, 2014). The test offers age-normed scores for *cognitive, language,* and *motor skills*, and a composite score *developmental status* of these three scales ($M = 100$, $SD = 15$); as well as normed scores for the subscales of receptive and expressive language and of fine and gross motor skills ($M = 10$, $SD = 3$). The norms of the German version of the BSID III are based on 878 German-speaking, monolingual children. The German version shows questionable (fine motor scale, $\alpha = .68$) to good internal reliability (gross motor scale, $\alpha = .83$). Content, construct and concurrent validity were established (Reuner & Rosenkranz, 2014). No Swiss German norms are available.

*Active vocabulary* was rated at $t_2$ and $t_3$ with the checklist "Language Assessment-Brief" (SBE-2-KT and SBE-3-KT; von Suchodoletz, Sachse, Kademann, & Tippelt, 2012). This parental questionnaire concerns active vocabulary in the children's primary language, consisting of 57 words at 24 months and 82 words at 36 months. The questionnaire shows excellent internal reliability ($\alpha = .98$, in both versions). Concurrent validity was established. Normed scores for German-speaking children are available for age groups ranging from 21 to 24 months and from 32 to 40 months, respectively. Since several children were outside this age range, percentages and not the standardized scores were used in the current analyses.

At $t_3$, *intelligence* was tested using the Snijders-Oomen Non-verbal Intelligence Test (SON-R 2½–7; Tellegen, Laros, & Petermann, 2007) at $t_3$. This test offers age-normed scores ($M = 100$, $SD = 15$) for intelligence, a sub-score for reasoning tests (categories, analogies, and situations), and a sub-score for performance tests (mosaics, puzzles, and patterns). The German norms of the SON-R 2½–7 are based on the weighted data from 1027 children, in which 18.1% of the children had an immigrant background. The overall scale shows excellent internal consistency ($\alpha = .90$). Content, concurrent and discriminant validity is established.

*2.4.2.2. Developmental dysfunction.* At $t_2$ and $t_3$, parents rated their children's *problem behavior* on the DSM-IV-oriented scales of the preschool version of the Child Behavior Check List (CBCL 1½–5; Achenbach & Rescorla, 2000): 45 problem-items are rated on a scale ranging from 0 (*not true (as far as you know)*) to 1 (*somewhat or sometimes true*) to 2 (*very true or often true*), and result in standardized *T*-scores ($M = 50$, $SD = 10$) in five subscales: affective problems (e.g., "underactive, slow moving, or lacks energy"), anxiety problems (e.g., "doesn't want to go out of home"), pervasive developmental problems (e.g., "disturbed by any change in routine"), attention deficit/hyperactivity problems (e.g., "quickly shifts from one activity to another"), and oppositional/defiant problems (e.g., "disobedient"). Normed scores are based on a sample of 700 U.S. children. German norms are not available. However, an examination of the U.S. standardization with a small German sample indicates that the U.S. norms can be applied in Germany without the risk of major bias (Elting, 2003), with the study showing comparable reliability ($\alpha > .70$) and concurrent validity to the U.S. version.

At $t_3$, *effortful control* was assessed with two tasks from a larger battery (Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996). The complete battery shows good internal consistency ($\alpha > .70$) and corresponds with maternal ratings of impulsivity and inhibitory control. Both tasks assess the children's executive processing of affective stimuli and the ability to suppress a dominant response and instead execute an alternative, subdominant response (Kochanska, Murray, & Harlan, 2000). First, in the "dinky toys" task the children can choose one of five attractive toys from a box. However, children are asked to put their hands on their knees, remain immobile, and only verbally name the preferred toy once the box is opened. Children's behavior is coded as 0% (grabs a toy), 20% (touches a toy), 40% (points to toy), 60% (removes hands from knees), 80% (hands twitching but remain on the knees), and 100% (hands rest immobile on the knees). This task was performed twice in a row. Second, an adapted version of the "gift delay" task was employed (Mulder, Hoofs, Verhagen, van der Veen, & Leseman, 2014). An attractively wrapped gift was presented to the children. They were instructed not to touch the gift until the researcher permitted them to do so after one minute. Children's behavior was coded as 0% (unwraps gift completely), 25% (tearing the wrapping paper), 50% (grabs the gift), 75% (touching the gift), and 100% (not touching the gift). This second task was performed only once at the end of the visit. Additionally, a composite score of the two tasks was calculated.

### 2.5. Analyses

#### 2.5.1. Attrition and missing data

Fig. 1 shows the flowchart of the participants' progress through the phases of the study. Prior to $t_1$, 15 of the 248 families at baseline dropped out of the program (5.6%), 11 families completed at least one measurement (i.e., interview or testing) of one wave in addition to baseline (4.4%), 15 families two waves (6.0%), and 208 families completed all waves (83.9%). The rates of overall attrition (14.9%) and differential attrition (5.4%) at $t_3$ meet the evidence standards of the What Works Clearinghouse™ (2017). We analyzed predictors of attrition at $t_3$ using binary logistic regression with the socio-demographics presented in Table 1 and the program site as predictors. Attrition was not statistically significantly affected by group membership, Odds Ratio (OR) = 2.09, 95% confidence interval (95% CI) [.83, 1.39], $p = .118$. However, results show that families with the following characteristics were more likely to drop out prior to completion of the program: fewer children (i.e., number of siblings), OR = .44, 95% CI [.18, 1.11], $p = .083$; recruited through additional measures, OR = .18, 95% CI [.06, .54], $p = .002$; younger mothers (i.e., mother's age on giving birth in years), OR = .89, 95% CI [.81, .97], $p = .007$; and mothers with post-compulsory education, OR = 3.90, 95% CI [1.32, 11.52], $p = .014$. No other statistically significant attrition effects with $p < .10$ were found.

Table 4 shows the amount of missing data in the single measures at $t_3$. There were two instruments with a large proportion of missing data: the SON-R 2½–7 and the effortful control tasks. The effect of missingness in these variables was analyzed based on the manifest data for the BSID III composite score. This score correlates statistically significantly with the SON-R 2½–7, $r = .58$, $p < .001$, and the combined effortful control tasks, $r = .47$, $p < .001$, there-

fore giving a meaningful criterion. Group (CG, IG) × Missingness (non-missing, missing) univariate ANOVAs were conducted on the scores of the BSID III at $t_3$. In the case of the SON-R 2½–7, there was no statistically significant effect of missingness, $F(1, 208) = 1.75$, $p = .155$, or of the interaction of Group and Missingness, $F(1, 208) = .09$, $p = .768$. In the case of the combined effortful control tasks, the interaction of Group and Missingness was not statistically significant, $F(1, 208) = .13$, $p = .716$. However, there was a statistically significant main effect of missingness, $F(1, 208) = 6.91$, $p = .009$. The group with missing values in the effortful control tasks showed a backlog corresponding to an effect size of .34. This difference suggests that the manifest data might misrepresent the true but unknown performance of the complete baseline sample. Provided that those participants with missing values potentially demonstrate the same performance in specific outcomes as others with similar known characteristics (i.e., MAR-assumption; Enders, 2010), the performance of the complete sample can be estimated—in our main analyses by applying multiple imputation (Little, 1988; Rubin, 1987). A sequential regression approach with predictive mean matching and a constant 25 iterations was carried out to estimate 40 complete data sets, including the randomization variables, the baseline variables presented in Table 1, predictors of attrition, and all dependent variables. Reported pooled results correspond to the means of the results of the 40 data sets.

#### 2.5.2. Data analysis

All analyses were performed with SPSS 24. Longitudinal data were analyzed using Generalized Estimating Equations (GEE; Twisk, 2013; Zeger & Liang, 1986). GEE employs quasi-likelihood to iteratively estimate marginal regression coefficients, and takes into account within-subject correlation between measurements points in longitudinal data. An autoregressive correlation was chosen to fit the data. Predictors were experimental group (control = reference group), measurement point (i.e., time), and Group × Time. Time was modeled as a scaled variable ranging from –2 ($t_1$) to 0 ($t_3$). Thus, the unstandardized regression coefficient for the experimental group represents the expected group difference at $t_3$ based on the linear regression. Glass's Δ was used to estimate effect sizes. This procedure is based on the standard deviation from the control group, and is therefore untainted by the intervention effects, making it more reflective of the population standard deviation (Glass, McGaw, & Smith, 1981).

Measures with only one measurement point (i.e., SON-R 2½–7, effortful control) were analyzed using linear regression. Single items were not imputed, but analyzed with the manifest data using binary logistic regression, *t*-tests, or $\chi^2$-tests, depending on the outcome. Statistically significant intervention effects and interactions of Group × Time were analyzed further. Time, risk-level and Group × Risk-level linear GEEs were performed on child outcomes in order to analyze differential effects depending on risk-level (i.e., low-risk, medium-risk, high-risk according to the HBS). If not otherwise stated one-tailed statistical testing was used. One-tailed tests provide more power to detect an effect in an a-priori defined direction (Cohen, Manion, & Morrison, 2011). Kimmel (1957) has formulated three criteria which justify the directional testing of hypotheses. These standards are met in the current study: First, the control group received no treatment, thus "a difference in the unpredicted direction, while possible, would be psychologically meaningless" (pp. 352–353). Second, equal to null results, negative results would speak against the effectiveness of PAT, thus "results in the unpredicted direction will, under no conditions, be used to determine a course of behavior different in any way from that determined by no difference at all" (Kimmel, 1957, p. 353). Third, previous trials show no or positive effects of PAT on child development (Neuhauser, 2014). Furthermore, theory unequivocally states that an enriched environment improves child outcomes and that especially high-risk families can be supported in the provision of a healthy environment (Bronfenbrenner & Morris, 2006; Rutter, 2011; Sameroff, 2010), thus "a directional hypothesis is deducible from psychological theory but results in the opposite direction are not deducible from coexisting psychological theory" (Kimmel, 1957, p. 353).

#### 2.5.3. Covariates

Children's age at testing in days, children's gender, global risk HBS (if not an explicit factor), and the mothers' German proficiency were included as covariates. The latter takes account of the fact that many instruments were either translated by intercultural interpreters, or were completed in a foreign language by mothers with little or no German proficiency, or may not have been precisely grasped by mothers with just sufficient German proficiency to complete the German versions of the instruments.

## 3. Results

The following sections report the intervention effects on child outcomes. Correlations between the measures are presented in Table 2. Children's manifest scores over the course of the three years are shown in Table 3. Table 4 shows the results of the GEEs and regressions.

### 3.1. Health

No statistically significant group differences were found concerning health. At 36 months, similar proportions of children were prescribed glasses (IG, 2.9%, CG, 1.0%), a hearing aid (IG, 3.8%, CG, 1.0%), a therapy (IG, 3.7%, CG, 7.7%), had completed the recommended vaccinations (IG, 98.1%, CG, 96.0%), and had never been hospitalized over the course of the three years (IG, 68.2%, CG, 73.6%). Two children needed hospitalization after an accident. The main reasons for hospitalization were acute infections (e.g., acute upper respiratory infections) or other isolated incidents, such as a tonsil operation or cyst surgery.

### 3.2. Adaptive behavior

Children from the IG complied with a statistically significant larger proportion of developmental milestones and self-help skills (see Table 4). Table 5 shows the percentage of children who complied with the single items of the checklist. Logistic regression analyses based on the manifest data showed significant effects in the case of sleeping alone, OR = 1.88, 95% CI [1.05, 3.36], $p = .017$, and scissor use, OR = 3.50, 95% CI [1.70, 7.18], $p = .001$. As a qualification, a ceiling effect can be observed for several items with compliances of nearly 100 percent (e.g., only five children in total did not use the word "no"). Adaptive behavior did not significantly correlate with global risk HBS (see Table 2). No effects of either risk-level nor Group × Risk-level were found in the GEEs.

### 3.3. Developmental competence

There was a significant intervention effect on the developmental status of the BSID III in the imputed results, and a marginally significant effect in the manifest results (see Table 4). Statistically, the intervention showed no effect on cognition and motor skills in the pooled results of the imputed data. However, there was a significant effect on cognition in the manifest results. The IG scored significantly higher in expressive language skills. The similar effect was observed in vocabulary as rated by the parents. Groups did not differ significantly in receptive language skills, but group differences declined statistically significantly over time, as demonstrated by

**Table 2**
Bivariate correlations between developmental outcomes at $t_3$.

| Outcome | 0 | 1 | 2 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Global risk HBS | – | | | | | | | | | | | | | |
| 1 Milestones | −.13 | – | | | | | | | | | | | | |
| 2 Self-help | −.09 | .19* | – | | | | | | | | | | | |
| 4 Cognition | −.36*** | .55*** | .16* | – | | | | | | | | | | |
| 5 Motor | −.23** | .52*** | .15* | .56*** | – | | | | | | | | | |
| 6 Language | −.36*** | .63*** | .19* | .69*** | .56*** | – | | | | | | | | |
| 7 Vocabulary | −.09 | .59*** | .19* | .44*** | .24** | .55*** | – | | | | | | | |
| 8 Intelligence | −.36*** | .44*** | .09 | .55*** | .46*** | .48*** | .31*** | – | | | | | | |
| 9 Affective | .07 | −.13 | .03 | −.14 | .04 | −.17* | −.19* | −.12 | – | | | | | |
| 10 Anxiety | .08 | −.22* | −.02 | −.14 | −.01 | −.18* | −.27** | −.16 | .58*** | – | | | | |
| 11 Developmental | .06 | −.19* | −.13 | −.20* | −.04 | −.26** | −.28** | −.13 | .71*** | .59*** | – | | | |
| 12 ADHD | .05 | −.08 | −.08 | −.06 | .09 | −.14 | −.17* | −.08 | .39*** | .28** | .37*** | – | | |
| 13 Oppositional | −.01 | −.09 | −.01 | .00 | .03 | .03 | −.12 | .04 | .39*** | .28** | .34*** | .38*** | – | |
| 14 Dinky toys | .13 | .27** | .19* | .43*** | .23** | .35*** | .17 | .21* | .06 | .03 | .07 | .03 | .11 | – |
| 15 Gift delay | .13 | .18* | .02 | .23** | .13 | .27** | .16 | . 24** | −.09 | −.02 | −.12 | −.03 | .01 | .13 |

* $p < .05$.
** $p < .01$.
*** $p < .001$, two-tailed.

**Table 3**
Manifest means and standard deviations of developmental outcomes.

| | | IG | | CG | | Low-risk | | Medium-risk | | High-risk | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Outcome | | *M* | (*SD*) | *M* | (*SD*) | *M* | (*SD*) | *M* | (*SD*) | *M* | (*SD*) |
| Self-help skills | $t_1$ | 52.7 | (25.9) | 43.2 | (24.4) | 47.9 | (24.7) | 49.2 | (25.9) | 47.0 | (26.1) |
| | $t_2$ | 56.0 | (15.8) | 48.2 | (15.0) | 54.8 | (15.7) | 50.7 | (16.5) | 52.2 | (15.1) |
| | $t_3$ | 61.3 | (21.8) | 54.9 | (22.3) | 56.8 | (21.8) | 60.1 | (23.4) | 56.5 | (20.8) |
| Milestones | $t_1$ | 88.6 | (10.1) | 87.0 | (10.9) | 89.2 | (10.3) | 87.4 | (10.2) | 87.2 | (11.0) |
| | $t_2$ | 91.0 | (11.8) | 86.6 | (14.2) | 90.1 | (10.0) | 88.7 | (13.7) | 88.0 | (14.7) |
| | $t_3$ | 83.3 | (13.9) | 80.1 | (15.4) | 83.3 | (14.5) | 80.9 | (15.8) | 81.5 | (13.1) |
| Cognitive skills | $t_1$ | 103.6 | (11.9) | 100.9 | (14.3) | 103.1 | (11.8) | 103.9 | (13.0) | 99.6 | (14.2) |
| | $t_2$ | 97.5 | (15.4) | 95.3 | (14.4) | 102.0 | (15.5) | 97.3 | (13.1) | 90.7 | (15.0) |
| | $t_3$ | 89.6 | (11.8) | 87.0 | (11.4) | 93.4 | (9.9) | 87.4 | (11.3) | 85.0 | (12.2) |
| Language skills | $t_1$ | 89.3 | (11.6) | 84.3 | (12.8) | 90.0 | (12.2) | 86.6 | (12.7) | 84.6 | (11.8) |
| | $t_2$ | 89.§ | (21.6) | 83.5 | (17.9) | 93.2 | (21.8) | 85.6 | (18.6) | 81.8 | (19.1) |
| | $t_3$ | 91.1 | (15.7) | 88.4 | (15.0) | 97.0 | (17.0) | 88.5 | (14.6) | 84.9 | (12.3) |
| Motor skills | $t_1$ | 99.1 | (13.4) | 97.6 | (15.0) | 98.8 | (14.2) | 98.3 | (14.4) | 98.2 | (13.9) |
| | $t_2$ | 102.3 | (15.7) | 101.4 | (14.5) | 103.9 | (12.9) | 102.7 | (15.9) | 98.9 | (15.6) |
| | $t_3$ | 99.1 | (12.6) | 96.1 | (13.0) | 99.7 | (12.8) | 98.7 | (12.3) | 94.0 | (13.1) |
| Vocabulary | $t_2$ | 59.1 | (30.7) | 48.9 | (30.1) | 60.1 | (32.6) | 52.2 | (30.8) | 52.4 | (28.8) |
| | $t_3$ | 81.7 | (19.7) | 73.9 | (25.8) | 79.4 | (24.8) | 75.4 | (24.9) | 80.2 | (18.4) |
| Intelligence | $t_3$ | 96.1 | (15.1) | 95.1 | (15.0) | 53.8 | (5.3) | 55.7 | (6.4) | 56.2 | (7.8) |
| Affective problems | $t_2$ | 55.5 | (6.8) | 55.2 | (6.5) | 54.7 | (6.5) | 56.4 | (7.1) | 56.4 | (8.7) |
| | $t_3$ | 54.8 | (5.9) | 57.0 | (8.6) | 53.0 | (4.1) | 55.3 | (6.6) | 55.2 | (6.7) |
| Anxiety problems | $t_2$ | 54.5 | (5.9) | 54.9 | (6.3) | 53.9 | (5.8) | 56.1 | (6.7) | 55.1 | (7.4) |
| | $t_3$ | 54.2 | (5.9) | 56.2 | (7.3) | 53.1 | (4.8) | 56.1 | (7.6) | 55.6 | (7.6) |
| Pervasive developmental problems | $t_2$ | 54.8 | (6.7) | 55.6 | (7.4) | 56.1 | (8.1) | 56.7 | (8.1) | 56.7 | (9.4) |
| | $t_3$ | 55.2 | (7.1) | 57.8 | (9.5) | 53.0 | (3.8) | 54.3 | (5.0) | 53.8 | (5.1) |
| ADHD problems | $t_2$ | 53.0 | (3.7) | 54.7 | (5.6) | 53.1 | (4.4) | 53.9 | (4.8) | 53.2 | (4.5) |
| | $t_3$ | 53.2 | (4.1) | 53.8 | (5.1) | 52.3 | (4.0) | 53.5 | (5.1) | 52.3 | (2.9) |
| Oppositional/defiant problems | $t_2$ | 53.2 | (4.4) | 52.4 | (4.1) | 52.9 | (5.3) | 53.8 | (4.9) | 52.3 | (4.4) |
| | $t_3$ | 53.3 | (5.5) | 52.9 | (4.2) | 53.8 | (5.3) | 55.7 | (6.4) | 56.2 | (7.8) |
| Dinky toys | $t_3$ | 31.2 | (33.3) | 22.0 | (25.3) | 38.5 | (29.9) | 24.5 | (29.1) | 18.9 | (28.3) |
| Gift delay | $t_3$ | 83.1 | (29.4) | 83.9 | (29.6) | 88.6 | (27.2) | 83.8 | (29.5) | 76.9 | (31.1) |

the negative interaction of Time × Group. Linear regression analyses with the imputed data for the single measurement points showed that adjusted group differences declined from $B = 1.20$, 95% CI [.52, 1.87], $p = .001$ at $t_1$, $B = .83$, 95% CI [.07, 1.59], $p = .160$, at $t_2$, to $B = .33$, 95% CI [−.33, 1.00], $p = .164$ at $t_3$. No intervention effects were found in the intelligence scores.

Global risk correlated significantly with all subscales of the BSID III (see Table 2). Separate GEEs showed that risk-level had a significant effect on the developmental status, $B = -3.23$, 95% CI [−5.06, −1.40], $p = .001$, cognition, $B = -3.46$, 95% CI [−5.72, −1.20], $p = .002$, receptive language, $B = -1.12$, 95% CI [−1.56, −.68], $p = .000$ and expressive language, $B = -.56$, 95% CI [−.94, −.17], $p = .003$. No significant interactions of Group and Risk-level were found.

### 3.4. Developmental dysfunction

Children from the IG showed statistically significantly less *problem behavior* as measured by the CBCL 1½–5. Specifically, there were significant effects on affective and pervasive developmental problems and on anxiety problems in the manifest data (see Table 4). By contrast, no statistically significant intervention effects were found on attention deficit/hyperactivity problems and on oppositional/defiant problems. There was a statistically significant Time × Group interaction in the affective problems scale. The groups did not differ at $t_2$, $B = -.09$, 95% CI [−1.93, 1.76], $p = .464$, but only at $t_3$, $B = -2.18$, 95% CI [−4.08, −.27], $p = .013$. No other statistically significant differences were observed. Finally, a significant effect was found in the manifest results of the dinky toys task of

**Table 4**
Intervention effects on child outcomes.

| | Manifest | | | | Pooled results | | | | | | Group |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Missing data at $t_3$ | Group | | | Group | | | Time × Group | | | at $t_3$ |
| Outcome | % | *B* (*SE*) | Wald | *p* | *B* | 95% CI | *p* | *B* | 95% CI | *p* | Glass Δ |
| **Adaptive behavior** | 18.4 | 5.79 (1.79) | 10.57 | .001 | 4.83 | [1.33, 8.34] | .004 | −.49 | [−2.94, 1.96] | .697 | .34 |
| Milestones | 18.4 | 4.70 (1.86) | 6.36 | .006 | 3.86 | [.18, 7.54] | .020 | .91 | [−1.24, 3.06] | .406 | .28 |
| Self-help skills | 18.4 | 6.85 (2.76) | 6.15 | .007 | 5.76 | [.16, 11.35] | .022 | −1.89 | [−6.26, 2.49] | .397 | .26 |
| **Developmental status** | 18.8 | 2.40 (1.47) | 2.68 | .051 | 2.20 | [−.20, 4.61] | .034 | −.47 | [−1.88, .94] | .518 | .21 |
| Cognition | 18.4 | 2.54 (1.52) | 2.80 | .047 | 2.20 | [−.65, 5.05] | .065 | −.33 | [−2.31, 1.65] | .745 | .19 |
| Motor | 18.8 | 1.68 (1.64) | 1.04 | .154 | 1.58 | [−1.20, 4.36] | .133 | .23 | [−1.73, 2.19] | .819 | .14 |
| Fine motor | 18.4 | .23 (.29) | .59 | .222 | 0.19 | [−.37, .76] | .253 | −.02 | [−.37, .33] | .895 | .09 |
| Gross motor | 18.8 | .29 (.30) | .92 | .169 | 0.28 | [−.28, .85] | .162 | .10 | [−.38, .57] | .696 | .08 |
| Language | 18.4 | 2.68 (2.00) | 1.78 | .091 | 2.78 | [−.57, 6.12] | .052 | −1.29 | [−3.21, .62] | .185 | .20 |
| Receptive language | 18.4 | .24 (.39) | .40 | .264 | 0.29 | [−.39, .98] | .201 | −.48 | [−.93, −.03] | .036 | .10 |
| Expressive language | 18.4 | .68 (.34) | 4.02 | .023 | 0.65 | [.07, 1.24] | .015 | .08 | [−.25, .41] | .632 | .28 |
| **Vocabulary** | 27.6 | 7.85 (3.15) | 6.19 | .007 | 8.15 | [1.49, 14.80] | .008 | −1.06 | [−8.61, 6.48] | .782 | .39 |
| **Problem behavior** | 26.4 | −1.49 (.65) | 5.29 | .011 | −1.27 | [−2.53, .00] | .025 | −.84 | [−2.10, .43] | .194 | −.30 |
| Affective | 26.4 | −2.38 (1.01) | 5.49 | .010 | −2.15 | [−4.10, −.21] | .015 | −2.16 | [−4.38, .06] | .056 | −.35 |
| Anxiety | 26.4 | −2.11 (.88) | 5.69 | .009 | −1.29 | [−3.25, .68] | .100 | −.97 | [−3.19, 1.25] | .390 | −.20 |
| Pervasive developmental | 26.4 | −2.49 (1.15) | 4.71 | .015 | −2.18 | [−4.52, .16] | .034 | −1.15 | [−3.62, 1.32] | .361 | −.29 |
| ADHD | 26.4 | −.77 (.65) | 2.18 | .070 | −0.79 | [−2.13, .56] | .125 | .65 | [−.90, 2.19] | .411 | −.18 |
| Oppositional/defiant | 26.4 | .20 (.69) | .09 | .385 | 0.05 | [−1.36, 1.46] | .473 | .41 | [−2.16, 1.06] | .505 | .01 |
| **Intelligence** | 31.4 | 2.69 (2.17) | 1.24 | .109 | 1.49 | [−1.69, 4.67] | .179 | - | - | - | .12 |
| Performance | 31.4 | 2.04 (1.98) | 1.03 | .153 | 1.36 | [−1.89, 4.61] | .206 | - | - | - | .11 |
| Reasoning | 31.4 | 3.09 (2.52) | 1.23 | .111 | 1.55 | [−2.30, 5.40] | .215 | - | - | - | .10 |
| **Effortful control** | 50.6 | .37 (3.61) | .10 | .459 | 3.40 | [−3.49, 10.28] | .166 | - | - | - | .15 |
| Dinky toys | 44.4 | 9.51 (4.76) | 2.00 | .024 | 6.46 | [−2.34, 15.26] | .075 | - | - | - | .21 |
| Gift delay | 42.5 | −.47 (4.95) | −.10 | .462 | 0.33 | [−9.61, 10.34] | .474 | - | - | - | .01 |

*Note.* Pooled data from 40 imputed datasets based on the randomization variables (age at randomization, program site, parent educator, single parents, twins, global risk HBS, mothers' German proficiency), predictors of attrition (number of siblings, mothers' age, recruitment with extra effort and education level), child characteristics (gender, gestational age, birth weight, first language, bilingualism), family characteristics (ISEI, mothers' life share in Switzerland), age at testing and all dependent variables.
GEE: time (−2 = $t_1$, −1 = $t_2$, 0 = $t_3$), group (0 = CG, 1 = IG), and Time × Group Generalized Linear Estimation Equation using an autoregressive correlation structure controlling for child's age at testing in days, child's gender, global risk HBS, and mothers' German proficiency. Two-tailed statistical testing was used for Time × Group interactions.
Linear regression analyses with the same covariates were used in the case of single measurement points; *t*-statistics are reported.
Effect size was calculated by dividing the adjusted mean difference by the *SD* of the CG.

**Table 5**
Adaptive behavior at 36 months.

| Outcome | IG % | CG % | low-risk % | medium-risk % | high-risk % |
|---|---|---|---|---|---|
| **Self-help skills** ($\alpha$ = .254) | | | | | |
| Sleeps alone | 52.8 | 39.6 | 45.8 | 52.2 | 37.9 |
| Needs baby bottle | 41.5 | 54.5 | 45.0 | 48.9 | 49.1 |
| Eats balanced and independently | 97.0 | 98.1 | 98.3 | 95.6 | 100.0 |
| Uses diapers at daytime | 39.0 | 40.2 | 36.7 | 44.4 | 35.1 |
| Uses diapers at nighttime | 65.4 | 66.3 | 76.7 | 59.6 | 64.4 |
| **Milestones** ($\alpha$ = .672) | | | | | |
| Stands on one leg | 92.2 | 94.9 | 89.3 | 96.5 | 93.2 |
| Walks stairs free-handedly | 96.3 | 95.0 | 93.3 | 97.8 | 95.0 |
| Rides a tricycle/ a like-a-bike | 92.5 | 94.1 | 98.3 | 92.2 | 89.8 |
| Uses scissors | 85.6 | 65.7 | 80.3 | 72.7 | 75.4 |
| Pretense play | 95.2 | 89.3 | 93.3 | 92.2 | 91.2 |
| Combines 3–5 words | 89.0 | 84.6 | 88.5 | 82.6 | 91.7 |
| Asks questions | 92.5 | 94.9 | 91.7 | 94.3 | 94.7 |
| Says "no" | 98.1 | 97.1 | 98.4 | 95.7 | 100.0 |
| Knows the colors | 59.4 | 49.0 | 70.0 | 47.1 | 49.2 |
| Says Ï" | 88.1 | 86.5 | 90.2 | 82.6 | 91.7 |
| Plays alone for 20 min. | 82.6 | 89.2 | 78.7 | 89.0 | 88.1 |
| Draws faces | 21.3 | 13.9 | 21.7 | 16.7 | 15.3 |
| Stops at the road before crossing | 92.8 | 91.2 | 92.9 | 93.8 | 88.5 |

the *effortful control* tasks. No group differences were found in the gift delay task.

As can be seen in Table 2, neither measure of developmental dysfunction correlated with the global risk score. Risk-level had a significant effect on the dinky toy task, $B = -11.01$, 95% CI [−21.55, −.47], $p = .021$. The interaction of Group and Risk-level was significant in the case of the total problem score, $B = -1.70$, 95% CI [−3.22, −.18], $p = .015$, affective problems, $B = -3.84$, 95% CI [−6.21, −1.46], $p = .001$, and in the manifest data of pervasive developmental problems, $B = -2.22$, 95% CI [−4.41, −.03], $p = .024$. Separate GEEs for the three risk-level groups showed that only in the high-risk group did group significantly predict the total problem score, $B = -2.19$, 95% CI [−4.35, −.04], $p = .023$, affective problems, $B = -5.08$, 95% CI [−8.64, −1.53], $p = .003$, and pervasive developmental problems, $B = -3.53$, 95% CI [−7.45, −.40], $p = .039$.

## 4. Discussion

This study indicates the effectiveness of the early intervention program PAT in the German-speaking part of Switzerland. The implementation of home-visiting programs to support psychosocially disadvantaged families during their children's first years is a recent development in German-speaking countries, and research on its effectiveness is sparse (Taubner et al., 2015), making the findings of the current study highly relevant. We found consistently positive effects on child outcomes (effect sizes ranging from −.01 to .39). In line with Bronfenbrenner's bioecological theory (Bronfenbrenner & Morris, 2006), the findings varied depending on the developmental outcome. Meaningful effect sizes above .20 (Cohen, 1988) were found in areas of developmental competence and developmental dysfunction. Children participating in PAT showed increased adaptive behavior, developmental status, language skills, and decreased problem behavior. By contrast, however, no meaningful increases were observed in children's health and cognitive or motor development. The study further showed that the intervention was not equally effective for all families

involved. In terms of developmental dysfunction, families exposed to high levels of psychosocial risk benefited most.

In general, results with and without imputation led to the same findings. However, only the manifest data led to statistically significant results in some outcomes. Therefore, relying on the pooled results is more conservative for most outcomes. Provided the MAR-assumption is valid, relying on the pooled results would compensate for participants with lower results leaving the study more often. Thus, the combination of the analyses ensures the credibility of the effects mentioned.

In line with previous findings, the intervention did not increase children's health or children's motor skills (Gomby, 2005). The overall healthy sample, with few health or other medical problems (e.g. only two hospital stays due to injuries) certainly accounts for this finding. Further, child access to the health care system is guaranteed in Switzerland and completely covered by the compulsory health insurance system, including nine pediatric screenings during the first three years after birth. Similarly, the study showed that over the course of the three years, children's motor development was within the normal range in the entire sample. The main reason for this finding is possibly that the acquisition of motor skills is largely dependent on child maturation (DiPietro, 2000; Laucht et al., 1997). Consequently, the promotion of motor skills is particularly important in interventions for children with disabilities (Farran, 2000), as opposed to when risks primarily stem from the environment. By contrast, PAT showed beneficial effects on children's adaptive behavior: we found small beneficial intervention effects both in self-help skills and in the developmental milestones (effect sizes = .26, .26). Children who took part in PAT were ahead of their peers without PAT in parent-led achievements such as to sleep alone (Sadeh, Tikotzky, & Scher, 2010).

As hypothesized, the intervention had positive effects on children's language skills, whereby it should be emphasized that actual language skills were most likely underestimated due to the fact that the BSID III is standardized for German-speaking children (Reuner & Rosenkranz, 2014). Moreover, translation of the test instructions was required for children not proficient in German. As a consequence, grammatical discrepancies between languages may have led to a cultural bias (cf. Goh et al., 2017). However, this effect was noticeable in both groups, therefore differences between the two groups nevertheless remain meaningful. In fact, the greatest effect size was found in children's expressive language skills (effect sizes = .28, .39). This finding underpins the strong influence of the environment on language development (Clark, 2016; Huttenlocher et al., 2010). In contrast to the stable beneficial effect of PAT on children's expressive language skills, the effect on children's receptive language skills faded over the course of the intervention. On the one hand, this finding corresponds to the hierarchical nature of language development and the different timetables of comprehension and production (Clark, 2016). Language comprehension builds both the foundation of and predicts language production (Fisher, 2017). On the other hand, previous findings have established an exact opposite trajectory in mutual influence (Raviv, Kessenich, & Morrison, 2004). Previous studies also have shown that children can compensate for an initial delay in language production (Hart & Risley, 1995). It thus remains to be seen in follow-up assessments whether this head start in language production was due to a short-term effective increase in verbal input, thus allowing the children from the control group to catch up in vocabulary, or whether this advantage has a long-term stability, including beyond the end of the intervention.

With regard to the cognitive measures, the study revealed conflicting findings. The positive effect in the manifest data of the BSID III subscale corresponds to the demonstrated influence of the environment on cognitive skills (Ayoub et al., 2009; Tucker-Drob & Harden, 2012). However, no group differences were found in the pooled results of the BSID III subscale or in the intelligence score of the SON-R 2½–7. It is noteworthy that while mean intelligence scores were within the normal range, children's cognitive skills were remarkably below the population mean. These discrepancies are probably due to differences between the instruments. The most prominent difference concerns the role of spoken language during testing. While the BSID III uses verbal instructions (which often had to be translated by our interpreters), the SON-R 2½–7 is explicitly non-verbal. Therefore, the SON-R 2½–7 is more appropriate in estimating the intelligence of children from immigrant backgrounds (Hagmann-von Arx, Meyer, & Grob, 2008). Given that there were a large proportion of children who primarily speak a foreign language in the present sample, the SON-R 2½–7 results may give a better approximation of children's true cognitive skills. Having said this, a predictive assessment of cognitive skills is only possible later in child development, i.e., from about four years onward (Shonkoff & Phillips, 2000). It will thus have to be left to future assessments to assess cognitive skills more reliably and to clarify whether the intervention has had an enduring positive effect.

Finally, PAT impacted outcomes of developmental dysfunction. Overall, PAT positively affected children's problem behavior (effect size = .30). Specifically, children participating in PAT showed less affective, anxiety (i.e., only in the manifest results) and pervasive developmental problems, and the group difference even increased over time. By contrast, no intervention effect was observed in externalizing problem behavior (i.e., ADHD and oppositional/defiant problems). Ultimately, we found a beneficial effect in on one of the effortful control tasks in the manifest data, but not in the imputed data. However, this finding needs to be interpreted with caution because there was a large proportion of missing data.

Increased psychosocial risk was associated with reduced developmental competence. This finding corroborates vast international evidence (Evans et al., 2013) and confirms the need for early support of at-risk families in Switzerland (Stamm et al., 2009). As far as our second hypothesis is concerned, high-risk families did not consistently benefit the most from PAT. The intervention affected developmental competence (i.e., language skills) irrespective of risk-level. By contrast, PAT affected developmental dysfunction (i.e., problem behavior) only in the most vulnerable families. In these families, the increase in problem behavior observed in the control group was not found in the intervention group. The bioecological model posits that proximal processes in disadvantaged environments are more directed towards dysfunction rather than towards competence (Bronfenbrenner & Morris, 2006). In high-risk families, effects should therefore only be expected in the case of dysfunction, but not in the case of competence. Contrary to this assumption, we found more favorable levels of both competence and dysfunction in high-risk families. Subgroup results should be interpreted with caution (Olds et al., 2007), and in this case are based on small sample sizes. Nevertheless, this discrepancy is consistent with previous findings and deserves further consideration. First and foremost, we found that group differences in dysfunction increased with age, but group differences in competence declined with age. This supports the finding that at earlier stages of development, parental investment mostly promotes competences, i.e., cognitive skills, while in older children parental investment mostly prevents dysfunction, i.e., problem behavior (Coneus et al., 2012). Previous studies also indicate that different qualities of parenting are relevant in these two developmental outcomes. Parental stimulation was related to competence (Tucker-Drob & Harden, 2012), while parental control was related to dysfunction (Karreman et al., 2006; Olson et al., 2002). Taken together, the current findings potentially suggest that the predominant reaction to dysfunction suggested by the bioecological model only occurs with the increasing age of the children and increasingly complex developmental or parenting tasks. Such a shift in focus, away from the promo-

tion of developmental competence towards the counteracting of developmental dysfunction, could potentially jeopardize the sustained effectiveness of PAT after the end of the program. The U.S. are increasingly focusing on more intensive programs that bridge the gap between early childhood and school entry (Reynolds, Magnuson, & Ou, 2010). This is needed because isolated early intervention measures alone often show little prospect of longer-term effects. Crucial for a sustainable effect is the domestic or school learning environment after the intervention. The current finding thus suggest that in order to support the sustained promotion of competences in high-risk families, support measures should be added to PAT that last into school age.

Taken together, the results of the present study suggest that PAT is an effective means to enhance child developmental outcomes in at-risk families. In many respects, this is remarkable. First, the effectiveness of home-visiting programs in a German-speaking country has not been proven before (Taubner et al., 2015). As a group, these countries differ in their social conditions—such as family poverty (OECD, 2018)—from the U.S., where the majority of previous studies have been conducted. Psychosocial risk is also lower in comparison to flagship programs such as the NFP program, which is specifically addressed to young, first-time mothers (Olds et al., 2007), who only made up a very small part of our own sample. Our results now indicate that the promotion of developmental competencies is equally effective in children growing up in more advantaged environments.

Second, we found larger effect sizes than previous PAT trials (Neuhauser, 2014). A possible explanation is that—corresponding to the more effective NFP program—our study involved PAT home visitors who were trained pediatric nurses, with additional training as parent counselors. Interventions with nurses as home visitors show the strongest effectiveness—probably because they are professionally competent in addressing parental concerns and are considered both credible and persuasive (Azzi-Lessing, 2011; Olds et al., 2007).

Third, while certainly limiting the validity of the results (see section below) a particular strength of the study is its unique sample. The majority of the participating families had a migration background and primarily spoke a foreign language. Undisputedly, these families are particularly vulnerable (Keels & Raver, 2009). The added complexity they represent in terms of program implementation and research (e.g. differing cultural backgrounds, need for translation) has meant that they have so far been neglected (Sama-Miller et al., 2017).

### 4.1. Limitation of findings

There are several limitations to the findings. The first set concerns the assessment and measuring instruments. While some outcomes were assessed using standardized instruments (e.g., BSID III), other outcomes were tested with non-standardized (i.e., adaptive behavior) and reduced (i.e., effortful control) instruments. It is uncertain whether these instruments are adequate to measure what they are intended to, so it remains to replicate the effects on these outcomes by applying standardized assessments. Also, in the case of outcomes derived from parental self-reporting (e.g., CBCL 1½–5, vocabulary) the bias cannot be ruled out that parents answered in a socially desirable manner. Finally, although the high proportion of migrant-background families was a particular feature of the study, since most measurements were language-dependent and required an interpreter/translation this potentially qualifies the validity of the results.

The second limitation concerns the age scope of the findings. Child outcomes were assessed up to the end of the intervention at 36 months of age. Various factors in child development are subject to continued change and are not necessarily stable (Belsky et al., 2007; Rutter, 2011). Even as the intervention progressed, group differences increasingly converged in certain developmental areas (e.g., language skills) at the same time as diverging in others (e.g., affective problems). These results underline the importance that Bronfenbrenner attributes to the time dimension (Bronfenbrenner & Morris, 2006) and the study of development over a long period of time. In short, follow-up assessments need to also examine the sustainability of the effects.

The third limitation concerns the scope of the findings. On the one hand, the successful and effective implementation of the program is not only promising in itself, but suggests that extending it as a service would have meaningful public advantages. On the other hand, the results fail to provide information about the underlying processes that mediate the positive findings. Although reference has been made to parental behavior in several contexts within this study, the assumption that the effects of PAT on child development are mediated by parent behavior remains theoretical (Peterson et al., 2007). The study itself cannot provide any data on this matter since the core processes of the bioecological model were not investigated. It would seem important, therefore, that future research focuses on the proximal processes in the expectation that this would elaborate on which parental behaviors are relevant for effective intervention.

Finally, despite the strength of the present findings in terms of the implementation of early intervention programs, specifically PAT, the study suffers from similar drawbacks to comparable trials: a small sample applied to a specific European region, with the sample decreasing over the course of the study. The findings are very promising, but require replication to reinforce their validity.

## 5. Conclusion

This RCT is the first to show the effectiveness of home-visiting as a preventive measure in at-risk families in the German-speaking part of Switzerland. After three years of intervention with the PAT program, children showed more age-appropriate adaptive behaviour, higher developmental competence regarding general developmental status and expressive language skills, and lower developmental dysfunction regarding problem behaviour. In terms of development competence, all families in this diverse sample benefited equally. On the other hand, effects on developmental dysfunction were only observed in high-risk families. These results are relevant beyond the Swiss context. They not only underline the effectiveness of the PAT program, but moreover demonstrate that home-visiting programs can also be used effectively and beneficially with families from culturally and linguistically diverse backgrounds.

## Declarations of interest

None.

## Acknowledgments

The findings reported here are based on research conducted as part of the “ZEPPELIN 0–3” research project, funded by the Swiss National Science Foundation [grant numbers 100019_134975, 100019_156610]. We are grateful to the Zurich Board of Education and especially the Parents as Teachers (PAT) staff for their collaboration and implementation of PAT. We thank the following organizations for their funding of the intervention: Ernst Göhner Foundation, Jacobs Foundation, Lottery Fund Canton of Zurich, Mercator Foundation, Paul Schiller Foundation, State Secretariat for Migration, and Vontobel Foundation. We gratefully acknowledge the contributions of the members of the research advisory board: Oskar Jenni, Daniel Schunk, Heidi Simoni, Margrit Stamm,

Ulrich Trautwein, and Ute Ziegenhain, as well as those of our former and current colleagues Ramona Eberli, Waltraud Sempert, and Franziska Templer. We thank the pediatricians from the University Children's Hospital Zurich and the student collaborators for their assistance in data collection, as well as Michael Bärtschi, Ute Bodmer, and Lukas Sigrist for their assistance in administration and coordination. We thank Thomas Skelton Robinson for proofreading. We gratefully acknowledge the valuable feedback on earlier versions of this manuscript provided by three anonymous reviewers. Our warmest thanks go to all the participating families.

## References


Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Ayoub, C., O'Connor, E., Rappolt-Schlictmann, G., Vallotton, C., Raikes, H., & Chazan-Cohen, R. (2009). Cognitive skill performance among young children living in poverty: Risk, change, and the promotive effects of early head start. *Early Childhood Research Quarterly*, *24*(3), 289–305. http://dx.doi.org/10.1016/j.ecresq.2009.04.001

Azzi-Lessing, L. (2011). Home visitation programs: Critical issues and future directions. *Early Childhood Research Quarterly*, *26*(4), 387–398. http://dx.doi.org/10.1016/j.ecresq.2011.03.005

Bayley, N. (2006). *Bayley scales of infant and toddler development* (3rd ed.). San Antonio, ZX: Harcourt Assessment.

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., & Owen, M. T. (2007). Are there long-term effects of early child care? *Child Development*, *78*(2), 681–701. http://dx.doi.org/10.1111/j.1467-8624.2007.01021.x

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology*, *22*(6), 723–742. http://dx.doi.org/10.1037/0012-1649.22.6.723

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner, W. Damon, R. M. Lerner, & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (Vol. 1) (6th ed., Vol. 1, pp. 793–828). Hoboken, NJ: John Wiley & Sons Inc.

Burger, K. (2013). *Early childhood care and education and equality of opportunity. Theoretical and empirical perspectives on social challenges*. Wiesbaden: Springer.

Burger, K., & Walk, M. (2016). Can children break the cycle of disadvantage? Structure and agency in the transmission of education across generations. *Social Psychology of Education*, *19*(4), 695–713. http://dx.doi.org/10.1007/s11218-016-9361-y

Cierpka, M., & Evers, O. (2015). Implementation and efficacy of early-childhood interventions in German-speaking countries. *Mental Health & Prevention*, *3*(3), 67–68. http://dx.doi.org/10.1016/j.mhp.2015.07.005

Clark, E. V. (2016). *First language acquisition* (3rd ed.). New York, NY: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). New York: Routledge.

Coneus, K., Laucht, M., & Reuss, K. (2012). The role of parental investments for cognitive and noncognitive skill formation—Evidence for the first 11 years of life. *Economics and Human Biology*, *10*(2), 189–209. http://dx.doi.org/10.1016/j.ehb.2011.01.003

Crul, M., & Vermeulen, H. (2006). Immigration, education and the Turkish second generation in five European nations: A comparative study. In C. A. Parsons, & T. M. Smeeding (Eds.), *Immigration and the transformation of Europe* (pp. 235–250). New York, NY: Cambridge University Press.

Diez Grieser, M. T., & Simoni, H. (2012). *Daten und Fakten zur Basisevaluation des Pogramms schritt:weise in der Deutschschweiz [Facts and figures on the basic evaluation of the program schritt:weise in German-speaking Switzerland]* Retrieved from. Zurich, Switzerland: Marie Meierhofer Institut für das Kind. https://www.a-primo.ch/download/pictures/d3/lc5g8a47y45twtm2ywb221wohdb04j/gesamtbericht_basisevaluation_mmi_kompr.pdf

DiPietro, J. A. (2000). Baby and the brain: Advances in child development. *Annual Review of Public Health*, *21*, 455–471.

Duncan, G. J., & Brooks-Gunn, J. (Eds.). (1997). *Consequences of growing up poor*. New York, NY: Russell Sage Foundation.

Elting, P. (2003). *Überprüfung der psychometrischen Parameter von CBCL 1 1/2–5 und C-TRF an einer deutschen Stichprobe [Assessment of the psychometric parameters of CBCL 1 1/2–5 and C-TRF in a German sample] (Doctoral dissertation)* Retrieved from Katalog der Deutschen Nationalbibliothek. (972529713/34).

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Evans, G. W., Li, D., & Whipple, S. S. (2013). Cumulative risk and child development. *Psychological Bulletin*, *139*(6), 1342–1396. http://dx.doi.org/10.1037/a0031808

Farran, D. C. (2000). Another decade of intervention for children who are low income or disabled: What do we know now? In J. P. Shonkoff, S. J. Meisels, J. P. Shonkoff, & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (pp. 510–548). New York, NY: Cambridge University Press.

Fisher, E. L. (2017). A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *Journal of Speech, Language, and Hearing Research*, *60*(10), 2935–2948. http://dx.doi.org/10.1044/2017_JSLHR-L-16-0310

Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research: A Quarterly Journal of Social Science Methodology and Quantitative Research*, *21*(1), 1–56. http://dx.doi.org/10.1016/0049-089X(92)90017-B

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Thousand Oaks, CA: Sage.

Goh, S. K. Y., Tham, E. K. H., Magiati, I., Sim, L., Sanmugam, S., Qiu, A., . . . & Rifkin-Graboi, A. (2017). Analysis of item-level bias in the Bayley-III language subscales: The validity and utility of standardized language assessment in a multilingual setting. *Journal of Speech, Language, and Hearing Research*, *60*(9), 2663–2671. http://dx.doi.org/10.1044/2017_JSLHR-L-16-0196

Gomby, D. S. (2005). *Home visitation in 2005: Outcomes for children and parents. Invest in kids working paper no. 7*. Washington, DC: Committee for Economic Development.

Gomby, D. S., Culross, P. L., & Behrman, R. E. (1999). Home visiting: Recent program evaluations—Analysis and recommendations. *Future of Children*, *9*(1), 4–26.

Hagmann-von Arx, P., Meyer, C. S., & Grob, A. (2008). Intelligenz- und Entwicklungsdiagnostik im deutschen Sprachraum [Intelligence and developmental scales in German-speaking countries]. *Kindheit und Entwicklung: Zeitschrift für Klinische Kinderpsychologie*, *17*(4), 232–242. http://dx.doi.org/10.1026/0942-5403.17.4.232

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes Publishing Company.

Hertzman, C., & Boyce, T. (2010). How experience gets under the skin to create gradients in developmental health. *Annual Review of Public Health*, *31*, 329. http://dx.doi.org/10.1146/annurev.publhealth.012809.103538

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365. http://dx.doi.org/10.1016/j.cogpsych.2010.08.002

Johnson, S., Moore, T., & Marlow, N. (2014). Using the Bayley-III to assess neurodevelopmental delay: Which cut-off should be used? *Pediatric Research*, *75*(5), 670–674. http://dx.doi.org/10.1038/pr.2014.10

Jungmann, T., Brand, T., Dähne, V., Herrmann, P., Günay, H., Sandner, M., . . . & Sierau, S. (2015). Comprehensive evaluation of the Pro Kind home visiting program: A summary of results. *Mental Health & Prevention*, *3*(3), 89–97. http://dx.doi.org/10.1016/j.mhp.2015.06.001

Karoly, L. A., Kilburn, M. R., & Cannon, J. S. (2005). *Early childhood interventions: Proven results, future promise*. Santa Monica, CA: RAND Corporation.

Karreman, A., van Tuijl, C., van Aken, M. A. G., & Dekovic, M. (2006). Parenting and self-regulation in preschoolers: A meta-analysis. *Infant and Child Development*, *15*(6), 561–579.

Keels, M., & Raver, C. C. (2009). Early learning experiences and outcomes for children of US immigrant families: Introduction to the special issue. *Early Childhood Research Quarterly*, *24*(4), 363–366. http://dx.doi.org/10.1016/j.ecresq.2009.09.002

Kimmel, H. D. (1957). Three criteria for the use of one-tailed tests. *Psychological Bulletin*, *54*(4), 351–353. http://dx.doi.org/10.1037/h0046737

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, *36*(2), 220–232. http://dx.doi.org/10.1037/0012-1649.36.2.220

Kochanska, G., Murray, K. T., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development*, *67*(2), 490–507. http://dx.doi.org/10.2307/1131828

Lanfranchi, A., & Neuhauser, A. (2013). ZEPPELIN 0-3: Theoretische Grundlagen, Konzept und Implementation des frühkindlichen Förderprogramms PAT—Mit Eltern Lernen [ZEPPELIN 0-3: Theoretical basis, concept, and implementation of a program for promoting early child development with PAT—learning with parents]. *Frühe Bildung*, *2*(1), 3–11. https://doi.org/10.1026/2191-9186/a000071

Laucht, M., Esser, G., & Schmidt, M. H. (1997). Developmental outcome of infants born with biological and psychosocial risks. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(7), 843–853. http://dx.doi.org/10.1111/j.1469-7610.1997.tb01602.x

Little, R. J. A. (1988). Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics*, *6*(3), 287–296. http://dx.doi.org/10.2307/1391878

Mulder, H., Hoofs, H., Verhagen, J., van der Veen, I., & Leseman, P. P. M. (2014). Psychometric properties and convergent and predictive validity of an executive function test battery for two-year-olds. *Frontiers in Psychology*, *5*, 733. http://dx.doi.org/10.3389/fpsyg.2014.00733

Neuhauser, A. (2014). A closer look at the effectiveness of early childhood education in at-risk families. *Mental Health and Prevention*, *2*(3–4), 43–57. http://dx.doi.org/10.1016/j.mhp.2014.09.002

Neuhauser, A., Ramseier, E., Schaub, S., Burkhardt, S. C. A., Templer, F., & Lanfranchi, A. (2015). Hard to reach families—A methodological approach to early recognition, recruitment, and randomization in an intervention study. *Mental Health & Prevention*, *3*(3), 79–88. https://do.org/10.1016/j.mhp.2015.07.002

OECD. (2016). *PISA 2015 results (volume I): Excellence and equity in education (summary)*. Paris: OECD Publishing.

OECD. (2018). *Poverty rate (indicator) (Publication no. 10.1787/0fe1315d-en)* Retrieved from. https://data.oecd.org/inequality/poverty-rate.htm

Olds, D. L. (2016). Building evidence to improve maternal and child health. *The Lancet, 38*, 105–107.
Olds, D. L., Sadler, L., & Kitzman, H. (2007). Programs for parents of infants and toddlers: Recent evidence from randomized trials. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *48*(3–4), 355–391. http://dx.doi.org/10.1111/j.1469-7610.2006.01702.x
Olson, S. L., Ceballo, R., & Park, C. (2002). Early problem behavior among children from low-income, mother-headed families: A multiple risk perspective. *Journal of Clinical Child and Adolescent Psychology*, *31*(4), 419–430. http://dx.doi.org/10.1207/153744202320802106
Peterson, C. A., Luze, G. J., Eshbaugh, E. M., Hyun-joo, J., & Kantz, K. R. (2007). Enhancing parent–child interactions through home visiting: Promising practice or unfulfilled promise? *Journal of Early Intervention*, *29*(2), 119–140.
Ramey, C. T., & Landesman Ramey, S. (1993). Home visiting programs and the health and development of young children. *The Future of Children*, *3*(3), 129–139. http://dx.doi.org/10.2307/1602546
Ramey, C. T., Sparling, J. J., & Landesman Ramey, S. (2012). *Abedarian. The ideas, the approach, and the findings*. Los Altos, CA: Sociometrics.
Raviv, T., Kessenich, M., & Morrison, F. J. (2004). A mediational model of the association between socioeconomic status and three-year-old language abilities: The role of parenting factors. *Early Childhood Research Quarterly*, *19*(4), 528–547. http://dx.doi.org/10.1016/j.ecresq.2004.10.007
Rayce, S. B., Rasmussen, I. S., Klest, S. K., Patras, J., & Pontoppidan, M. (2017). Effects of parenting interventions for at-risk parents with infants: A systematic review and meta-analyses. *BMJ Open*, *7*(12), e015707 http://dx.doi.org/10.1136/bmjopen-2016-015707
Reuner, G., & Rosenkranz, J. (2014). *Bayley scales of infant and toddler development, third edition—Deutsche Version [German version]*. Frankfurt am Main: Pearson Assessment.
Reynolds, A. J., Magnuson, K. A., & Ou, S. R. (2010). Preschool-to-third grade programs and practices: A review of research. *Children & Youth Services Review*, *32*(8), 1121–1131. http://dx.doi.org/10.1016/j.childyouth.2009.10.017
Robling, M., Bekkers, M.-J., Bell, K., Butler, C. C., Cannings-John, R., Channon, S., . . . & Torgerson, D. (2016). Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (building blocks): A pragmatic randomised controlled trial. *The Lancet*, *387*(10014), 146–155. http://dx.doi.org/10.1016/S0140-6736(15)00392-X
Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons, Inc.
Rutter, M. (2000). Psychosocial influences: Critiques, findings, and research needs. *Development and Psychopathology*, *12*(3), 375–405. http://dx.doi.org/10.1017/S0954579400003072
Rutter, M. (2011). Biological and experiential influences on psychological development. In D. P. Keating, & D. P. Keating (Eds.), *Nature and nurture in early child development* (pp. 7–44). New York, NY: Cambridge University Press.
Sadeh, A., Tikotzky, L., & Scher, A. (2010). Parenting and infant sleep. *Sleep Medicine Reviews*, *14*(2), 89–96. http://dx.doi.org/10.1016/j.smrv.2009.05.003
Sama-Miller, E., Akers, L., Mraz-Esposito, A., Zukiewicz, M., Avellar, S., Paulsell, D., . . . & Del Grosso, P. (2017). *Home visiting evidence of effectiveness review: Executive summary*. Washington DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
Sameroff, A. J. (2010). A unified theory of development: A dialectic integration of nature and nurture. *Child Development*, *81*(1), 6–22. http://dx.doi.org/10.1111/j.1467-8624.2009.01378.x
Schoon, I. (2006). *Risk and resilience: Adaptations in changing times*. New York, NY: Cambridge University Press.
Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The HighScope Perry Preschool study through age 40*. Ypsilanti, MI: HighScope Press.
Shonkoff, J. P. (2010). Building a new biodevelopmental framework to guide the future of early childhood policy. *Child Development*, *81*(1), 357–367. http://dx.doi.org/10.1111/j.1467-8624.2009.01399.x
Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.
Sidor, A., Eickhorst, A., Stasch, M., & Cierpka, M. (2012). Einschätzung der Risikobelastung in Familien im Rahmen von Frühen Hilfen: Die Heidelberger Belastungsskala (HBS) und ihre Gütekriterien [Assessing the risk exposure in families within the scope of early intervention: The Heidelberger Belastungsskala (HBS, Heidelberg Stress Scale) and its quality criteria]. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, *61*(10), 766–780, doi:20.500.11780/3681.
Sierau, S., Dähne, V., Brand, T., Kurtz, V., von Klitzing, K., & Jungmann, T. (2016). Effects of home visitation on maternal competencies, family environment, and child development: A randomized controlled trial. *Prevention Science: The Official Journal of The Society for Prevention Research*, *17*(1), 40–51. http://dx.doi.org/10.1007/s11121-015-0573-8
Sroufe, L. A., Egeland, B., Carlson, E. A., & Collins, W. A. (2005). *The development of the person: The Minnesota study of risk and adaptation from birth to adulthood*. New York, NY: Guilford Publications.
Stamm, M., Reinwand, V. I., Burger, K., Schmid, K., Viehhauser, M., & Muheim, V. (2009). *Frühkindliche Bildung in der Schweiz. Eine Grundlagenstudie im Auftrag der UNESCO-Kommission Schweiz [Early childhood education in Switzerland. A study on behalf of the Swiss Commission for UNESCO]* Retrieved from. Fribourg: University of Fribourg, Switzerland. http://www.unesco.ch/wp-content/uploads/2017/03/Grundlagenstudie.pdf
Sweet, M. A., & Appelbaum, M. I. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development*, *75*(5), 1435–1456. http://dx.doi.org/10.1111/j.1467-8624.2004.00750.x
Tassé, M. J. (2009). Adaptive behavior assessment and the diagnosis of mental retardation in capital cases. *Applied Neuropsychology*, *16*(2), 114–123. http://dx.doi.org/10.1080/09084280902864451
Taubner, S., Wolter, S., & Rabung, S. (2015). Effectiveness of early-intervention programs in German-speaking countries—A meta-analysis. *Mental Health & Prevention*, *3*(3), 69–78. http://dx.doi.org/10.1016/j.mhp.2015.07.001
Tellegen, P. J., Laros, J. A., & Petermann, F. (2007). *Snijders-Oomen non-verbaler Intelligenztest 2 1/2–7—Revidierte Fassung [Snijders-Oomen nonverbal intelligence test. SON-R 2 1/2–7—Revised version]*. Göttingen: Hogrefe.
Tucker-Drob, E. M., & Harden, K. P. (2012). Early childhood cognitive development and parental cognitive stimulation: Evidence for reciprocal gene-environment transactions. *Developmental Science*, *15*(2), 250–259.
Twisk, J. W. R. (2013). *Applied longitudinal data analysis for epidemiology: A practical guide*. New York, NY: Cambridge University Press.
von Suchodoletz, W., Sachse, S., Kademann, S., & Tippelt, S. (2012). *Früherkennung von Sprachentwicklungsstörungen: Der SBE-2-KT und SBE-3-KT für zwei- bzw. dreijährige Kinder [Early diagnosis of language development disorders: The SBE-2-KT and SBE-3-KT for children aged 2 or 3 years]*. Stuttgart: Kohlhammer.
Werner, E. E., & Smith, R. S. (2001). *Journeys from childhood to midlife: Risk, resilience, and recovery*. Ithaca, NY: Cornell University Press.
What Works Clearinghouse™. (2017). *Standards handbook (version 4.0)* Retrieved from. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf
Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*(1), 121–130. http://dx.doi.org/10.2307/2531248