

RESEARCH ARTICLE

WILEY

A meta-analysis of the effects of after-school programmes among youth with marginalized identities

Kirsten M. Christensen¹ | Kristen P. Kremer² |
Cyanea Y. S. Poon¹ | Jean E. Rhodes¹

¹Department of Psychology, University of Massachusetts Boston, Boston, Massachusetts, USA

²Department of Sociology, Anthropology, and Social Work, Kansas State University, Manhattan, Kansas, USA

Correspondence

Kirsten M. Christensen, Department of Psychology, University of Massachusetts Boston, 100 Morrissey Blvd, Boston, MA 02125, USA.

Email: kirsten.christens001@umb.edu

Abstract

After-school programmes (ASPs) often provide youth with a safer alternative to unstructured time while providing a context for building skills and forging positive relationships with programme staff and peers. ASPs may be particularly effective for youth with marginalized identities, including youth of colour and youth from low-income backgrounds. Despite this promise, few rigorous evaluations of ASPs have been conducted and even fewer meta-analyses have investigated the effects of ASPs among youth with marginalized identities. Using a multi-level meta-analysis of 615 effect sizes across 56 studies (overall $n = 128,538$), the current study examined the overall effects of ASPs on internalizing, externalizing, school-related, social functioning, and self-perception/identity outcomes among kindergarten through 12th grade youth with marginalized identities. Results indicated ASPs to have a small, yet significant positive overall effect on youth outcomes ($g = 0.2049$, $p = .001$, 95% CI = 0.08–0.33). Moderator analyses revealed significant differences in effects based on outcome source and outcome measure type. Given the ubiquity of ASPs and the challenges that youth experiencing marginalization face, this study uniquely adds to the existing literature and outlines important implications and recommendations for research, policy, and practice.

KEYWORDS

after-school programmes, marginalized, meta-analysis, systematic review, youth

1 | INTRODUCTION

After-school programmes (ASPs) can be critical contexts for positive development, affording millions of youth in the United States opportunities to access developmental assets, build skills, and connect with peers and caring adults. Most ASPs operate during the school year to provide academic enrichment and recreational activities, often with diverse goals. Research has most often examined academic achievement (e.g., Lauer et al., 2006) and behavioural outcomes (e.g., Taheri & Welsh, 2016) associated with engagement in ASPs, although some studies also suggest potential gains in psychosocial and identity-related outcomes for some youth (e.g., Durlak, Weissberg, & Pachan, 2010). In fact, it is this wide array of goals and outcomes that has historically attracted funding and support for after-school programming. These programmes have often been developed to support youth with marginalized identities including youth of colour, youth from low-socioeconomic status backgrounds, or other characteristics that put them 'at risk' for negative outcomes, such as youth attending lower performing/Title I schools, youth with a history of trauma, or youth who engage in risky behaviours like substance use or delinquency.

Despite the growing consensus regarding ASP benefits and overall acceptance of these programmes as a positive force in the lives of youth with marginalized identities, the research base on ASPs is remarkably limited. Most research in the after-school field has included a relatively small number of participants, has not been based on strong experimental methods (i.e., primarily correlational in nature), and/or has focused on a limited range of outcomes. Furthermore, little is known about the specific youth and programme characteristics that may be associated with increased ASP effectiveness. As such, this study sought to systematically examine the current evidence to understand the overall effectiveness of ASPs for youth with marginalized identities.

1.1 | Outcomes associated with ASPs for youth with marginalized identities

Youth in low-income communities may be at risk for more negative developmental trajectories due to the environmental circumstances and stressors they often face. Structural and systemic injustice in particular increase the likelihood of these youth experiencing more trauma, stress, and marginalization (e.g., racism-related stress, exposure to community violence; Halpern, 2005) than their more privileged peer counterparts (Forrest-Bank, Nicotera, Bassett, & Ferrarone, 2016; Leventhal & Brooks-Gunn, 2000). Unstructured, unsupervised time between 3 and 6 p.m. may also increase opportunities to engage in unsafe behaviours, such as substance use or risky sexual behaviours (Afterschool Alliance, 2014b). Thus, when accessible, ASPs often provide structured spaces for youth to not only avoid risk but to develop skills and receive social support and connectedness.

Although limited, research demonstrates that ASPs may be particularly beneficial for youth with marginalized identities, providing skills and assets that not only prevent engagement in risky behaviour but promote healthy development (Hirsch, 2005). For example, ASPs, even those not exclusively academic-focused, can offer resources that compensate for under-resourced school systems or overworked parents by providing skills-building activities, homework help, and tutoring. Such programmes have been associated with positive academic outcomes, including school bonding and retention among Latinx (Diaz, 2005) and African American (Dotterer, McHale, & Crouter, 2007) adolescents, as well as higher academic adjustment for youth of all races (Marsh & Kleitman, 2002). Moreover, lower-income youths' participation in ASPs is associated with academic benefits (Miller, 2003). Among low-income

youth of colour, for example, participation in ASPs is associated with increased academic achievement and lower school dropout rates (Vandell, Reisner, & Pierce, 2007).

There are also improved psychological and identity-related outcomes associated with marginalized, low-income, and youth of colour's participation in organized programmes (Pedersen & Seidman, 2005), such as increases in self-esteem, ethnic identity (Riggs, Bohnert, Guzman, & Davidson, 2010), emotional regulation, interpersonal and social behaviour, peer acceptance, and psychological adjustment (Durlak & Weissberg, 2007; Hirsch, 2005). Furthermore, youth may use organized programmes as places to safely explore their identities in ways that counter potentially negative peer contexts (Hirsch, 2005). Moreover, youth involvement in ASPs is associated with decreased externalizing behaviours such as drinking, substance use, and antisocial behaviour (Pedersen & Seidman, 2005) among low-income adolescents and/or youth of colour.

Finally, correlational research suggests that youth engagement in ASPs is linked with positive mental health outcomes, especially for youth from low-income families (Marshall et al., 1997) and urban communities (Frazier, Cappella, & Atkins, 2007). This is important given that the prevalence of anxiety and depressive disorders is higher in youth of colour as well as youth living in economically disadvantaged communities (Van Voorhees et al., 2008). Although no randomized controlled studies have examined the effects of ASPs on mental health outcomes, quasi-experimental studies of arts-based programmes have demonstrated promise for positive mental health outcomes for youth with marginalized backgrounds, including reductions in depression and anxiety (e.g., Wright et al., 2006). Case study and qualitative designs also contribute to theories that arts programmes may increase the development of protective factors such as coping skills (Rhodes & Schechter, 2014) and self-regulation (Averett, Crowe, & Hall, 2015).

1.2 | Limitations in ASP research with marginalized youth populations

Although ASPs have widely been accepted as promising contexts for positive development, especially among youth with marginalized identities, a number of important limitations exist in the literature. For example, it is noteworthy that research on the developmental benefits of ASP participation is somewhat mixed. Although many have been positive, some investigations of ASPs, including 21st Century Community Learning Centers (CCLCs; e.g., James-Burdumy, Dynarski, & Deke, 2007), have found null and sometimes even harmful effects, particularly for sports involvement (Eccles, Barber, Stone, & Hunt, 2003) or when youth engage in unusually high levels of extracurriculars (i.e., the 'overscheduling hypothesis'; Marsh & Kleitman, 2002). Although most studies examining the 'overscheduling hypothesis' apply to youth of higher socio-economic status whose families have the means to enrol them in multiple programmes, youth with marginalized backgrounds may still experience this phenomenon if they engage in ASPs in conjunction with athletics or other responsibilities such as time-intensive after-school jobs (Hirschman & Voloshin, 2007) or caring for younger siblings (Halpern, 2005). Furthermore, although engagement in highly structured after-school activities is associated with lower levels of antisocial behaviour, participation in programmes with low structure may be associated with more deviant behaviour as a result of learned negative behaviours from peers and lack of support from programme staff (Mahoney & Stattin, 2000).

The lack of attention to the differential quality and resources of after-school programming available to marginalized youth relative to their more privileged peers may help to explain some of the mixed effects in the literature. Research on ASPs, as well as school environments and general programme evaluation, suggests a few quality indicators for promoting positive youth outcomes, such as staffing ratios, staff qualifications and training, staff turnover, racial and social climate, and effective leadership. These elements are important regardless of whether youth are engaging in recreational activities, developing skills, or doing academic work (e.g., Astroth, Garza, & Taylor, 2004; Simpkins, Riggs, Ngo, Vest Ettekal, & Okamoto, 2017). These quality indicators are particularly relevant for marginalized youth who, during the school day, are likely to receive educational programming that is less well-resourced and more subject to poor quality factors than privileged youth (e.g., EdBuild, 2019). After-school programming is subject to the same systems and limitations present in school settings that contribute to substantial achievement gaps.

These limitations include poorer funding, higher teacher–student ratio, workforce issues (e.g., high turnover, low pay, poor staff training), undue disciplinary emphasis, the likelihood of racially disproportionate discipline practices, insufficient physical activity time, and an overly academic structure in younger grades (e.g., EdBuild, 2019). Thus, ASPs have the potential to parallel or reproduce these inequities, varying in quality in proportion to the marginalization of the youth served. Indeed, most attention and funding are often focused on increasing the number of ASP slots available, rather than on efforts to improve programme quality (Hynes & Sanders, 2011). Overall, the existing literature on after-school programming lacks a critical examination of how programme quality might affect youth outcomes, especially for youth from marginalized backgrounds or communities. Consequently, the current meta-analysis is a start at examining proxies of some of these quality indicators. This represents an extension of past research, which has been primarily focused on comparisons of youth enrolled in ASP versus those in control or comparison groups, without considering key quality indicators.

As a result of the limitations above, existing ASP evaluations vary so extensively that it is difficult to determine their true overall effects. To this end, meta-analysis provides a rigorous method to systematically examine the effect of interventions on outcomes. Meta-analysis combines the results of multiple individual evaluations in ways that can produce more reliable and precise impact estimates and permit comparisons across approaches and characteristics.

1.3 | Prior meta-analyses of ASP effectiveness

To date, seven meta-analyses (Ciocanel, Power, Eriksen, & Gillings, 2017; Durlak et al., 2010; Kremer, Maynard, Polanin, Vaughn, & Sarteschi, 2015; Lauer et al., 2006; Lester, Chow, & Melton, 2020; Taheri & Welsh, 2016; Zief, Lauver, & Maynard, 2006) have been conducted on the effects of ASPs. Collectively, they have yielded mixed findings regarding the academic, behavioural, and psychosocial benefits of ASPs. Key outcomes previously examined include primarily academic achievement and other school-related behaviours such as school bonding, attendance, and academic adjustment, as well as externalizing and behavioural outcomes such as drug use, sexual health outcomes, problem behaviours, and positive social behaviours. The effect sizes of these outcomes across studies have ranged from negative $d = -0.05$ to moderate 0.50, with most falling in the small effect size range according to Cohen's d (1998).

One more recent, high-quality meta-analysis conducted by Kremer et al. (2015) examined the effects of ASPs on externalizing behaviours and school attendance among primary and secondary school-aged 'at-risk' youth. The study sought to redress many of the limitations of previous meta-analyses by following rigorous systematic review guidelines and conduction and reporting standards of meta-analyses set by the Campbell Collaboration. As a result, the current study draws on and extends Kremer et al. (2015) protocol. Findings from the meta-analysis of 24 studies, including 109,282 youth, revealed non-significant effects for attendance and externalizing behaviours, and none of the tested moderator variables (i.e., study design, grade level, contact, control group, programme type, and focus) explained the variance between studies. However, no moderator variables related to programme quality were assessed (e.g., implementation fidelity, staff training), and the examined outcomes (i.e., externalizing and attendance) were limited. Given that ASP quality may vary by privilege and access to resources, marginalized youth may receive lower-quality ASPs, suggesting the importance of considering these additional moderating factors.

More generally, past efforts to comprehensively evaluate the effects of ASPs, particularly among youth facing a number of environmental stressors and sociodemographic risk factors, include important limitations. To begin, the methods utilized in prior meta-analyses of the effects of ASPs have varied significantly in terms of quality, as well as the extent to which researchers have explicitly outlined the specific methods employed. Although establishing methods a priori for meta-analyses decreases the likelihood of selection bias and error and increases the reliability and validity of the results, many reviews do not specify their inclusion criteria, literature search processes, and data extraction and coding methods. Relatedly, many meta-analyses have not accounted for risk of other types of bias.

Assessing these risks is critical for understanding the internal validity of included studies, and thus, how the results should be interpreted.

Moreover, few meta-analyses have measured the effects of ASPs on a full range of youth outcomes. Rather, most prior meta-analyses have focused on one or a few important outcomes, often related to academics or behaviour. In particular, very few meta-analyses to date have examined general psychosocial outcomes such as psychological distress and self-perceptions (Ciocanel et al., 2017), and even fewer have examined specific internalizing mental health symptoms, as well as identity development outcomes. These developmental outcomes are deserving investigation, given that childhood and adolescence are critical stages of identity formation and youth are at increased risk for mental health issues (Office of Adolescent Health, 2019).

Finally, a number of moderating factors with the potential to influence the magnitude of an ASP's effectiveness have not yet been extensively examined. For example, variability in the youth population served, activities engaged in, and other programme variables such as youth outcome category, youth characteristics, programme characteristics, implementation fidelity, and methodological characteristics as contributing factors to positive youth outcomes. Although Kremer et al. (2015) redressed many of these limitations, they focused on a particular band of outcomes. Thus, the current meta-analysis aimed to update and replicate the rigorous methods implemented in Kremer et al. (2015), coding for similar studies and moderators but including a wider range of outcomes that ASPs are often thought to influence.

1.4 | Potential moderators of ASP effectiveness

1.4.1 | Youth outcome category

Although ASPs have historically been thought to promote a wide array of positive outcomes for youth, meta-analyses on the effectiveness of ASPs have remained relatively limited in their scope of assessed outcomes. In particular, outcomes such as mental health or internalizing symptoms as well as self-perception and identity-related outcomes have been understudied. As a result, the current study examined not only the overall effectiveness of ASPs on youth outcomes as a whole, but also specific outcome category (i.e., school, social, internalizing, externalizing, and self-perception/identity) as a potential moderating variable of programme effectiveness.

1.4.2 | Youth sociodemographic characteristics

Youth characteristics have been found to moderate the effectiveness of ASPs. Youths' age (often by grade level) is associated with ASP participation rates and some outcomes (Riggs & Greenberg, 2004), since younger youth may also be easier to recruit than older youth due to increased levels of maturity and independence among older adolescents (Hofferth & Jankuniene, 2001). Furthermore, youth who have academic challenges may have more difficulty meeting their peers' academic level later in school, even if they participate in ASPs with academic components (Bowman, Donovan, & Burns, 2000). Selection biases can also arise in relation to behavioural outcomes if older youth with pre-existing challenges are referred to ASPs by parents or teachers to increase structured, supervised time, engendering age-related differences in youth outcomes (Vandell & Posner, 1999).

Also important to consider is youth gender, ethnic/racial identity, and socio-economic status. Although girls and boys participate in ASPs at about the same rate (Afterschool Alliance, 2014a), there is limited existing research on the role of youth gender on outcomes. Research investigating other organized youth programmes such as mentoring suggests that males may benefit more than females (Raposa et al., 2019), although more research is necessary to explore whether similar trends occur in ASPs. Moreover, ASPs have been found to be beneficial for youth of colour and lower-income youth. For example, there are increased psychological and identity-related outcomes associated

with low-income and/or youths of colour participation in organized programmes (Pedersen & Seidman, 2005) such as self-esteem, ethnic identity (Riggs et al., 2010), emotional regulation, interpersonal and social behaviour, peer acceptance, and psychological adjustment (Durlak & Weissberg, 2007; Hirsch, 2005). In addition, participation in after-school/out-of-school programming is associated with increased academic achievement and lower school drop-out rates for low-income youth of colour (Vandell et al., 2007).

1.4.3 | Programme characteristics

A number of characteristics related to ASPs themselves are important to consider as moderating factors of their effectiveness on youth outcomes. For example, the amount of contact that youth have with the ASP has the potential to moderate programme effectiveness. However, the research is mixed, with some studies suggesting the amount of time youth spend in an ASP is not necessarily important (Roth, Malone, & Brooks-Gunn, 2010), and others suggesting more benefit with more regular contact or exposure (McComb & Scott-Little, 2003).

Additionally, not all ASPs have the explicit purpose of promoting positive youth outcomes. While some programmes may target specific youth challenges or outcomes, such as academic-focused programmes, other programmes may focus on more unstructured, recreational activities. However, programmes that are more recreational may still be associated with positive outcomes in some domains, even if they are not intentionally targeting specific outcomes. As non-experimental research (i.e., qualitative, correlational) suggests that youths' level of engagement, interest, and motivation in their programme is influenced by the types of activities completed, with some research pointing to higher engagement in sports and creative arts activities than academic-related activities (Shernoff & Vandell, 2007), it is important to consider how programme focus may influence ASP effectiveness.

Furthermore, programme type (i.e., local/regional, national) may also influence the effectiveness of ASPs on youth outcomes. For example, national programmes may be more widely recognized and thus have more stable and available funding, resources, and/or structure, whereas local programmes may have more difficulty accessing evaluation services or financial resources to conduct self-evaluations to assess their quality and influence on youth (Scott-Little, Hamann, & Jurs, 2002). On the other hand, local grassroots programmes may be inclined to prioritize focus on programme youth and community needs, whereas national programmes may privilege the implementation of uniform national requirements. Given the variability in ASPs, exploring programme type can serve as an exploratory proxy for factors related to 'how' programmes might be running.

Location (i.e., urban, suburban, rural) may also influence programmes' resources or implementation, as youth in rural and urban communities may face challenges outside of programme spaces that can influence ASP's effectiveness on youth outcomes. For example, cities and rural regions have higher poverty rates than suburban geographical areas in the United States (Burton, Mattingly, Pedroza, & Welsh, 2017). In addition, youth in low-income, urban communities may be at a particular risk for more negative developmental trajectories as a result of the sociohistorical, systemic, and structural oppression concentrated in these areas. This includes experiencing more traumas and marginalization than their more privileged peer counterparts (Forrest-Bank et al., 2016; Leventhal & Brooks-Gunn, 2000), due to racism-related stress, structural economic disadvantage, or an unjust and violent criminal justice system (Halpern, 2005). As youth living in urban contexts who are identified as 'at-risk' are most frequently prioritized by funders and policymakers in after-school programming, it is important to consider how programme location may be associated with youth outcomes.

1.4.4 | Implementation fidelity and staff training

Implementation fidelity is the implementation of 'strategies that monitor and enhance the accuracy and consistency of an intervention to ensure it is implemented as planned and that each component is delivered in a comparable

manner to all study participants over time' (Smith, Daunic, & Taylor, 2007, p. 121). Assessing implementation fidelity is important in ASPs as it helps explain whether outcomes are explained by the actual intervention and its effectiveness, or by other confounding factors (Maynard, Peters, Vaughn, & Sarteschi, 2013). Staff can also play a powerful role in ASPs and youth outcomes, providing a breadth and depth of support for youth with marginalized identities (Hirsch, 2005). Indeed, staff training and education are associated with youth outcomes such as school adjustment and may have indirect effects on other youth outcomes (Yohalem & Wilson-Ahlstrom, 2010). Thus, it is important to examine the association between staff training and education and youth outcomes.

1.4.5 | Methodological characteristics

In addition to youth- and programme-related characteristics, it is equally important to assess the methodological predictors of ASP effect sizes. These characteristics include publication type, as published studies typically report higher effect sizes than unpublished studies due to biases in the peer-review process that make it more likely for studies with statistically significant results to be published (Pigott, Valentine, Polanin, Williams, & Canada, 2013). Furthermore, research design characteristics, such as the type of comparison group (i.e., waitlist or no intervention, treatment as usual, alternative intervention or programme), and whether the study implemented a randomized control design or a quasi-experimental design are considered, as studies using quasi-experimental designs often yield greater effect sizes than randomized control designs (Cheung & Slavin, 2016). In ASP research, one factor is often self-selection bias, in which youth who engage in ASPs are also those who already have access to other resources that promote positive outcomes (Developmental Services Group, Inc., 2010). Lastly, outcome source (e.g., report by youth, parents, teachers, or official records) and outcome measure type (e.g., surveys/rating scales compared with archival reports/school records) may be important moderating variables of programme effectiveness. In addition to the potential for self-report bias, researcher-developed tests tend to yield larger effect sizes than standardized measures (Cheung & Slavin, 2016). Moreover, the narrower outcomes found in quantifiable archival/school reports (e.g., attendance, GPA) may not reflect as much nuance in the broader psychosocial gains youth may make in outcomes typically assessed in self-report scales and surveys. It is therefore important to examine these methodological characteristics.

2 | CURRENT STUDY

This systematic review and meta-analysis synthesized existing evidence on the overall effectiveness of ASPs on (1) mental health (i.e., internalizing), (2) behavioural (i.e., externalizing), (3) self-perception/identity, (4) social functioning, and (5) school-related outcomes among youth with marginalized identities. To date, few comprehensive meta-analyses have examined the effects of youths' participation in ASPs, and almost none include mental health outcomes in their assessments. Given the prevalence of youths' involvement in ASPs, the marginalization certain youth face, and the potential for these programmes to prevent risk and promote positive outcomes, this is an understudied topic. Although empirical evidence is available through individual studies of youth programmes, meta-analyses provide cumulative evidence of the effectiveness of these programmes. To this end, the current study brings the existing literature up to date by incorporating new studies published since 2014 and empirically assessing new youth outcomes with robust statistical methods. Unlike most previous reviews and meta-analyses, this study employed methodologies following rigorous guidelines established by the Campbell Collaboration (2014) to account for the multiple limitations in prior reviews. Further, the current study assessed youth and programme characteristics, implementation fidelity, staff training, and study design/methodological characteristics as moderators to determine whether these factors change the magnitude of the effect of ASPs. It was hypothesized that compared with youth in comparison conditions, youth participation in ASPs would be statistically significantly associated with

positive overall outcomes. Additionally, it was hypothesized that the effects of ASP participation would be moderated by some or all of the following: outcome category, youth characteristics, programme characteristics, and study design/methodological characteristics. Given the limited extant research and expectations that minimal data would be reported on implementation fidelity as well as staff training and education, these were investigated as exploratory potential moderators.

3 | METHOD

3.1 | Study selection

In this meta-analysis, ASPs were defined as organized programmes supervised by adults occurring outside of school hours during the regular school year. Programmes occurring only during the summer or during school hours were excluded due to the potential for summer camps or in-school curricula to have a different influence on youth (e.g., Bialeschki, Henderson, & James, 2007). To separate ASPs from other more limited after-school activities or sports-based activities and in accordance with the 21st Century Community Learning Center definition, an ASP must have offered more than one activity to youth to help them meet local or state standards of academic achievement and must have offered 'a broad array of enrichment activities (to) complement their regular academic programs' (U.S. Department of Education, 2014). Additionally, programmes that only included mentoring or tutoring were excluded from the current study, since these programmes do not meet the aforementioned definition and have also been examined as unique interventions for youth (Raposa et al., 2019). However, studies that included mentoring or tutoring as components of an ASP in addition to other activities were eligible for inclusion.

To build upon previous literature, the study inclusion criteria for the current study aligned with those established by Kremer et al. (2015), with the exception of outcome category. Studies that utilized experimental or quasi-experimental designs to examine the effects of ASPs on internalizing, externalizing, school-related, social functioning and self-perception/identity outcomes in primary and secondary school-aged youth were included in the current study. To be eligible, studies must have used a comparison group (i.e., wait list, no intervention, treatment as usual, or alternative intervention) and reported baseline measures of outcome variables. Internalizing outcomes were broadly defined to include mental-health-related outcomes, including mood or other psychiatric disorders or symptoms. Externalizing behaviour outcomes were broadly defined to include behavioural issues such as delinquency, disruptive behaviour, or substance use. School-related outcomes were broadly defined to include outcomes such as academic functioning/grades, school attendance, school engagement, disciplinary referrals, or teacher- and school-based peer relationships. Social-functioning-related outcomes were broadly defined to include supportive relationships, prosocial behaviour, social support, and interpersonal skills. Finally, self-perception outcomes were broadly defined to include self-esteem, self-concept, identity, confidence, self-efficacy, and belonging. Measures completed by youth, teachers, parents, or administrative data were eligible for inclusion in this study.

Eligible studies included youth participants in grades kindergarten through 12 who are considered marginalized based on meeting at least one of the following criteria: (1) performing below grade level or having low scores on measures of academic achievement; (2) attending low-performing or Title I schools; (3) having characteristics or identities that are often associated with risk for low academic achievement, such as low socio-economic status/receiving free or reduced lunch, racial or ethnic minority background (i.e., youth of colour), single parent family, limited English proficiency, or experienced trauma, abuse, or neglect; or (4) engaging in risky behaviours such as truancy/elopeing from home, substance use, or delinquency (adopted from Kremer et al., 2015). In order for the study as a whole to meet this criterion, at least 50% of the youth in the study sample must have identified with at least one of these characteristics. Samples including youth holding both marginalized and privileged identities (e.g., entire sample is from a Title 1 school and more than 50% of the sample identifies as White) were considered to meet inclusion criteria.

Multiple sources were utilized to identify published and unpublished works on the effects of youth involvement in ASPs. Meta-analyses in the current study included previously identified and included articles from Kremer et al. (2015). In addition, primary studies excluded by Kremer et al. (2015) as a result of not meeting outcome eligibility criteria were included in the current meta-analysis to analyse a broader range of previously unexamined outcomes. In addition to these articles, new empirical articles were searched for and systematically reviewed from the following electronic databases: Academic Search Premier, ERIC, ProQuest Dissertations and Theses, PsycINFO, Social Sciences Citation Index, Social Services Abstracts, Social Work Abstracts, and Sociological Abstracts. The search strategy used slight variations of the keywords 'after-school programs' AND (evaluation OR treatment OR intervention OR outcome) to obtain articles published since May 2014 (i.e., when Kremer et al., 2015 search was completed) through June 2019. During this search, reference lists of previous reviews and included articles were also searched to identify any other articles that met the inclusion criteria. A forward citation search of key papers was also conducted using Google Scholar to identify articles that have cited key ASP reviews and meta-analyses. In an effort to reduce publication bias, unpublished works (i.e., dissertations and theses) that met inclusion criteria were also included in the systematic review process and subsequent analyses.

The initial search process of publications, unpublished dissertations/theses, and reports from 2014 and later yielded 6,055 potential articles. This was in addition to the 2,587 records identified from Kremer et al. (2015). Duplicate studies were screened out prior to evaluation for inclusion. See Figure 1 for a description of the study selection process. Article titles and abstracts of studies obtained through the above search criteria were screened for relevance and preliminary inclusion by the first author and one research assistant. Any uncertainties about the relevance of articles at this phase were addressed by obtaining the full-text article to be briefly screened. Articles that appeared to be relevant and potentially eligible based on review of the title and abstract were obtained in full text and screened by the first author using the screening tool developed by Kremer et al. (2015). Subsequent to this initial screen, potentially eligible studies were further reviewed by the first author and a Bachelor's level research assistant to determine final inclusion. Any discrepancies were discussed and resolved between these two researchers. Fifty-six studies (representing 57 samples) met the final inclusion criteria for this meta-analysis.

3.2 | Study coding procedures

All included studies were coded using a coding instrument adapted from Kremer et al. (2015). The codebook contained five sections, including methodological/study characteristics, youth characteristics, programme characteristics, research methods and quality descriptors, and effect size data. The first author independently coded all data to calculate effect sizes for all study outcomes as well as all data relevant to the aforementioned moderator variables, and two research assistants double-coded all effect size and moderator data. Overall inter-rater agreement for moderator data across all studies was 90.76%. Overall inter-rater agreement for effect size data across all studies was 98.76%. Discrepancies in coding were discussed and resolved among all coders to achieve final consensus.

Youth outcome category included assessments of school/academic performance, social functioning, externalizing/behavioural symptoms, internalizing/mental health symptoms, and self-perception/identity. Youth characteristics included age (i.e., elementary, middle, high school, mixed), race/ethnicity, gender, and socio-economic status (i.e., income level/free-reduced lunch). Programme characteristics included location (i.e., urban, suburban, rural, mixed), type (i.e., local/regional, national), focus/content (i.e., academic, non-academic, mixed), and amount of contact (i.e., once per week or less, 3–4 times per week, daily). Implementation fidelity characteristics included programme operationalization (i.e., well-defined procedures, written manual), fidelity measurement (i.e., measurement of adherence to protocol), and frequency of fidelity measurement (i.e., weekly, monthly, quarterly, biannually, annually). Staff training was categorized by amount of implementer training, including no training (0), some training (1), and comprehensive training provided (2). Where available, staff level of education was also coded. Due to the expectation that minimal data would be reported on implementation fidelity, staff training, and education, these moderators

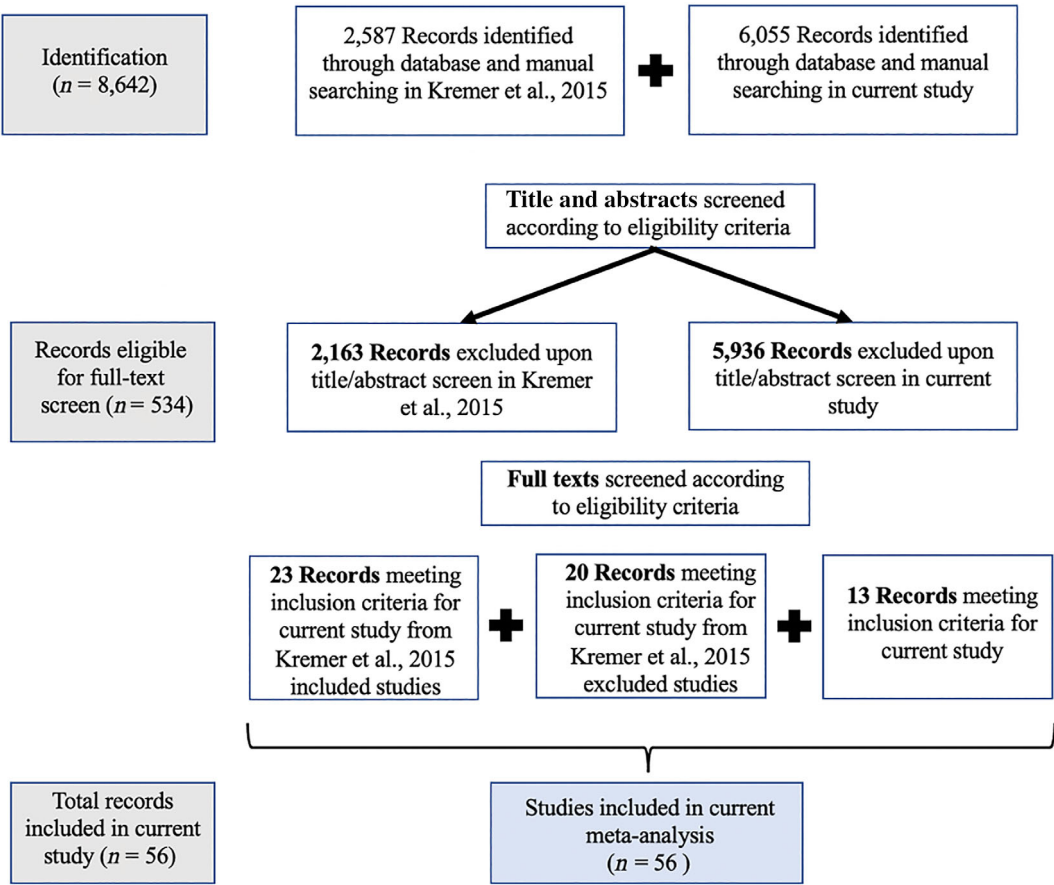


FIGURE 1 Study selection process flow chart

were considered exploratory. Finally, study design/methodological characteristics included publication type (i.e., published, unpublished), comparison type (i.e., waitlist/no intervention, TAU, alternative), design (i.e., RCT, quasi-experimental), study sample size, outcome source (i.e., self-report, parent-report, schoolteacher-report, official record, other report), and outcome measure type (i.e., archival/school record, rating scale/survey/questionnaire, behavioural observation). Official records refer to measures not rated by teachers, but typically on file with school administrative offices. Examples include attendance data, GPA, disciplinary history on school record, standardized test scores, report card grades, school enrollment, dropout, and school graduation/completion. Teacher-reported outcomes refer to measures specifically designated as teacher-rated scales (i.e., completed by schoolteachers, rather than administrative data). These measures often focused on teachers' perspectives of child competencies via teacher-rated emotional, behavioural, and academic measures. Some examples include the Richardson Teacher-Child Rating Scale, the Cowen Teacher-Child Rating Scale, and the Teacher Assessment of Student Skills.

3.3 | Calculation of effect sizes

Effect sizes were calculated through the standardized mean difference between treatment and control groups for each outcome, with a positive standardized mean value indicating better performance in the treatment group. Effect sizes (standardized mean differences) were then transformed using Hedge's *g* to correct for small sample size bias

and differences in sample sizes across studies. Hedge's g corrects for sample size to allow effect sizes with greater precision (i.e., effect sizes drawn from larger samples) to be assigned greater weight in analyses. In cases where data including sample sizes, means, standard deviations/standard errors, or other information necessary for effect size calculation was not reported in full-text articles, study authors were contacted via email in an effort to obtain missing information.

3.4 | Data analyses

All statistical analyses were conducted in R (version 4.0.3) using the *metafor* package (Assink & Wibbelink, 2016). A random-effects multi-level (i.e., three-level) meta-analysis was conducted to examine the effects of ASPs on overall youth outcomes. This analysis accounts for the presumed diversity in study populations, programmes, settings, and outcomes, and estimates three sources of variance—sampling variance of observed effect sizes, variance between effect sizes obtained from the same study, and variance between studies. Subsequently, heterogeneity tests were conducted to better understand variance within the summary statistics, including (1) Cochran's Q -test to determine whether sources of heterogeneity remain (i.e., indicates the presence versus the absence of heterogeneity), and (2) I^2 value to indicate the extent of the heterogeneity, or the proportion of observed heterogeneity that can be attributed to true heterogeneity, rather than sampling error (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006).

Moderator analyses were subsequently conducted as heterogeneity tests met the recommended threshold for analysis (i.e., I^2 values at least between 50% and 70%; Higgins & Thompson, 2002). Moderators consisted of those determined a priori based on previous research (Kremer et al., 2015), including type of outcome, youth characteristics, programme characteristics, implementation fidelity, staff training and education, and methodological characteristics.

3.5 | Analysis of publication and other bias

Bias can interfere with one's ability to draw conclusions about the results or effects gleaned from meta-analyses. Studies with low validity or a review including studies of varying levels of validity, for example, have the potential to yield biased estimates for outcomes, and thus, can lead to biased, skewed, or misinterpreted findings. As a result, all studies included in analyses were evaluated for five common threats to internal validity using the Cochrane Collaboration's tool for assessing bias: (1) selection bias, (2) performance bias, (3) detection bias, (4) attrition bias, and (5) reporting bias. All articles were independently (i.e., double) coded by the first author and a doctoral-level research assistant. To rate study bias in each of these domains (i.e., low, high, or unclear risk of bias), articles were examined and coded for random sequence generation, allocation concealment, blinding of participants, blinding of outcome assessment, incomplete outcome data, and selective reporting (Higgins & Green, 2011). Inter-rater agreement among the two coders was initially 76.76%. Upon collaboratively reviewing the Cochrane Collaboration risk of bias coding manual again, all discrepancies were discussed and resolved in order to reach 100% consensus on all codes. A risk-of-bias bar graph was produced to reflect the level of bias present in the included studies (see Figure 2).

Several analyses were also conducted to test for potential publication and other bias. First, a funnel plot analysis was conducted to test for publication bias (assessed through funnel symmetry or asymmetry; Crombie & Davies, 2009). A trim-and-fill analysis using aggregated data at the publication level (Duval & Tweedie, 2000; Viechtbauer, 2010) was then used to explore whether there were any missing effect sizes. In trim-and-fill analysis, missing effect sizes on the left side of the funnel plot indicate that non-significant or negative results are missing, whereas missing effect sizes on the right side of the plot suggest a selection bias, or over-representation of studies with larger effect sizes (i.e., studies including characteristics that make them systematically associated with larger effects).

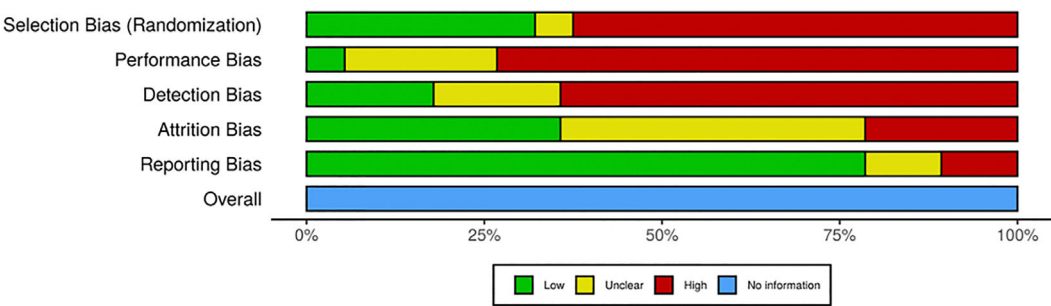


FIGURE 2 Risk of bias across studies

Next, sensitivity analyses were conducted to determine whether there were certain studies influencing the overall effects to be larger or smaller. According to the Cochrane Collaboration (Higgins & Green, 2011), issues warranting sensitivity analyses are almost always identified during the article coding process, where issues are first recognized. Given the wide range of sample sizes and robustness of effect sizes across included studies in the current meta-analysis, leave-one-out sensitivity analyses were conducted. The overall effect size of outcomes was recalculated 57 times, with each iteration removing one study from the total sample to examine the unique influence of each study on the overall effect size (Viechtbauer & Cheung, 2010).

4 | RESULTS

Descriptive information related to study and sample characteristics, intervention characteristics, and summarized study information can be found in Tables 1–3 respectively.

4.1 | Overall effect of ASPs on youth outcomes

The overall effect size across 56 studies and 615 effect sizes (128,538 youth participants) was $g = 0.2049$ ($p = .001$; 95% CI = 0.08–0.33), indicating a statistically significant small effect size (Cohen, 1998). Subsequent analyses revealed significant heterogeneity across studies (σ^2 level 3 = .17, $p < .001$) and within studies (σ^2 level 2 = .33, $p < .001$).

4.2 | Moderator analyses of ASP effectiveness

Given that tests of homogeneity indicated a substantial amount of heterogeneity ($Q = 12510.55$; $I^2 = 98.85\%$), follow-up moderator analyses were conducted to explore potential moderating variables of the association between ASP participation and overall youth outcomes. Results of moderator analyses on youth outcome category, youth characteristics, programme characteristics, implementation fidelity, staff training and education, and methodological characteristics are summarized in Table 4.

Results indicated that youth outcome category, youth characteristics, programme characteristics, implementation fidelity, and training characteristics were not significant moderators. Frequency of fidelity measurement (i.e., weekly, monthly, quarterly, biannually, annually) and staff level of education could not be examined as

TABLE 1 Study and sample characteristics ($N = 57$)

Characteristic	<i>n</i> (%)
Publication year	
1990–1999	8 (14.04%)
2000–2009	24 (42.11%)
2010–2019	25 (43.86%)
Grade level	
Elementary	21 (36.84%)
Middle	16 (28.07%)
High school	2 (3.51%)
Mixed	17 (29.82%)
Control group condition	
Nothing or waitlist	43 (75.44%)
Treatment as usual	8 (14.04%)
Alternative treatment	6 (10.53%)
Sample size	
0–150	20 (35.09%)
151–300	10 (17.54%)
301–600	10 (17.54%)
601 and greater	17 (29.82%)
Multiple marginalized characteristics	
Only one marginalized characteristic	10 (17.54%)
More than one marginalized characteristic	47 (82.46%)
Mean Age ^a	10.73
Free/reduced lunch ^b	76.23%
Percent male ^c	50.91%
Majority race in study sample ^d	
African American/Black	28 (49.12%)
Asian/Pacific Islander	2 (3.51%)
Hispanic/Latinx	7 (12.28%)
Indigenous/native American/Alaska native	1 (1.75%)
White	10 (17.54%)
Research design type	
Randomized controlled	23 (40.35%)
Quasi-experimental	34 (59.65%)
Publication type	
Journal article	33 (57.89%)
Government report	2 (3.51%)
Thesis or dissertation	14 (24.56%)
Other unpublished report	8 (14.04%)
Country	
USA	52 (91.23%)
UK	5 (8.77%)

Note: $N = 57$ as Biggart et al. (2014) includes two reports (elementary & middle school data).

^aIncludes 25 studies with age data.

^bIncludes 28 studies with FRL data.

^cIncludes 51 studies with gender data.

^dIncludes 48 studies that had race information.

TABLE 2 ASP intervention characteristics (N = 57)

Characteristic	n (%)
Setting	
School	21 (36.8%)
Community-based organization	29 (50.9%)
Mixed	3 (5.3%)
Unsure	4 (7.0%)
Programme focus	
Academic	11 (19.3%)
Non-academic	13 (22.8%)
Mixed	29 (50.9%)
Unsure	4 (7.0%)
Manual used for intervention	
No	15 (26.3%)
Yes, for the entire programme	22 (38.6%)
Yes, for part of the programme	16 (28.1%)
Unsure	4 (7.0%)
Programme coverage	
National	17 (29.8%)
Local	40 (70.2%)
Number of sessions	
0–50	16 (28.1%)
51–100	9 (15.8%)
101–150	6 (10.5%)
151 and greater	7 (12.3%)
Unsure	17 (29.8%)
Length of sessions	
Less than 1 h	2 (3.5%)
1–1.59 h	15 (26.3%)
2–2.59 h	14 (24.6%)
3–3.59 h	14 (24.6%)
4 h or more	3 (5.3%)
Unsure	9 (15.8%)
Weekly contact frequency	
Once	5 (8.8%)
Twice	9 (15.8%)
Three to four	19 (33.3%)
Five	13 (22.8%)
Unsure	10 (17.5%)
Ongoing staff supervision/coaching	
No	10 (17.5%)
Yes, supervision built into programme	13 (22.8%)
Some (unsystematic) oversight provided	9 (15.8%)
Unsure	25 (43.9%)

Note: N = 57 as Biggart et al. (2014) include two reports (elementary and middle school data).

TABLE 3 Summary of included studies (N = 57)

Study citation	Grade level ^a	Risk/marginalized characteristic ^b	Programme coverage	Programme location	Programme focus	Weekly contact	Total sample (N)	Study design ^c	Outcomes ^d
Arcaira et al. (2010)	2	1, 2, 3	National	Urban	Academic	-	870	QED	School
Avery (2013)	3	1, 2	Local	Urban	Academic	2 × week	238	RCT	School, SP
Baker and Witt (1996)	1	1, 2	Local	Urban	Mixed	1–4 × week	302	QED	School, SP
Belgrave et al. (2000)	4	1, 2, 4	Local	Urban	Non-academic	1 × week	184	QED	SP
Bergin et al. (1992)	1	1, 3, 2, 4, 7	Local	Urban	Academic	3–4 × week	24	QED	School
Biggart et al. (2013)	1	1, 3	National	Urban	Academic	3–4 × week	621	RCT	School
Biggart et al. (2014)	1	1, 3	National	Urban	Academic	3–4 × week	464	RCT	School
Blumer and Werner Wilson (2010)	4	3, 7	Local	-	Non-academic	3–4 × week	58	QED	Int
Brock et al. (2018)	1	1, 2, 3	Local	Urban	Mixed	3–4 × week	87	RCT	School, Ext
Carr (2015)	2	1, 2, 3	Local	Urban	Mixed	-	716	QED	School, Ext
Dreyer (2010)	4	1, 2	Local	Urban	Academic	Daily	135	QED	School
Foley and Eddins (2001)	1	1, 2	Local	Urban	Mixed	-	5,915	QED	School
Frazier et al. (2013)	4	1, 2, 4	Local	Urban	Non-academic	Daily	127	QED	School, social, Ext
Gottfredson et al. (2004)	4	2, 7	Local	Urban	Mixed	3–4 × week	801	RCT	Social, Ext
Gottfredson et al. (2010)	2	1, 2, 7	Local	-	Mixed	3–4 × week	447	QED	Ext
Graham et al. (2015)	1	1, 2, 3	Local	Urban	Mixed	3–4 × week	66	RCT	School, social, Ext, SP

TABLE 3 (Continued)

Study citation	Grade level ^a	Risk/marginalized characteristic ^b	Programme coverage	Programme location	Programme focus	Weekly contact	Total sample (N)	Study design ^c	Outcomes ^d
Grassi (2012)	4	1, 2	Local	Urban	Academic	3–4 × week	256	QED	School, social, SP
Green (2010)	2	1, 2, 3	National	Urban	Mixed	3–4 × week	73	QED	School
Grolnick, Farkas, Sohmer, Michaels, and Valsiner (2007)	2	1, 2	Local	Urban	Academic	3–4 × week	90	RCT	School, Ext, Int, SP
Helseth and Frazier (2018)	1	1, 2	Local	Urban	Non-academic	2 × week	61	QED	Social
Hirsch et al. (2011)	3	1, 2	Local	Urban	Mixed	3–4 × week	547	RCT	School, social, Ext, Int, SP
James-Burdumy et al. (2007)	1	1, 2, 3	National	Urban	–	Daily	2,288	RCT	School, social, Ext, SP
James-Burdumy et al. (2008)	2	1, 2, 3	National	Urban & Rural	–	3–4 × week	3,831	QED	School, Ext
Jenner and Jenner (2007)	1	1, 2, 3	National	Urban & Rural	Mixed	3–4 × week	1,192	QED	School
Kaufman et al. (2018)	2	2	Local	Rural	Non-academic	–	167	RCT	Ext, SP
La Torre et al. (2019)	1	1, 2, 3	Local	Urban	Mixed	Daily	5,618	QED	School
LaFrance et al. (2001)	4	1, 2, 7	Local	Urban	Mixed	Daily	242	QED	School, Ext
Langberg et al. (2007)	2	1, 2, 3, 7	Local	Urban	Academic	3–4 × week	48	RCT	School, SP
Langberg et al. (2008)	4	3	Local	Suburban	Mixed	2 × week	37	RCT	School
Le et al. (2011)	4	2	Local	Urban	Mixed	Daily	338	QED	School, Ext
Lecroy et al. (2018)	2	1, 2	Local	Urban	Non-academic	1 × week	801	RCT	School, Ext, SP
McKinney (1995)	1	2, 3	Local	Rural	Mixed	2 × week	44	RCT	School
Moldow (2007)	1	1, 2	Local	Urban	Mixed	Daily	365	QED	School

(Continues)

TABLE 3 (Continued)

Study citation	Grade level ^a	Risk/marginalized characteristic ^b	Programme coverage	Programme location	Programme focus	Weekly contact	Total sample (N)	Study design ^c	Outcomes ^d
Molina et al. (2008)	1	2, 7	National	Urban	Non-academic	2 × week	23	RCT	School, Ext, Int,
Nance et al. (2000)	4	3	Local	Urban	Mixed	-	278	QED	School
Nguyen (2007)	4	1, 3	National	-	Mixed	3-4 × week	26,330	QED	School, social, Ext
O'Hare et al. (2015)	1	1	Local	Urban	Non-academic	2 × week	589	RCT	School, social, Ext, Int
Oyserman et al. (2002)	2	1, 2	Local	Urban	Non-academic	1 × week	208	QED	School, Ext, SP
Pastchal-Temple (2012)	2	1, 3	National	Rural	Mixed	2 × week	66	QED	School, Ext
Pettway (2005)	2	1, 2, 3	Local	-	Academic	3-4 × week	167	QED	School
Phillips (1999)	4	1, 2	Local	Urban	Mixed	Daily	180	RCT	Social, Int, SP
Pierce and Shields (1998)	4	2, 6, 7	Local	Urban	Non-academic	1 × week	783	QED	School, social, Ext, Int, SP
Prenovost (2001)	2	1, 2, 3, 5	Local	Urban	Mixed	Daily	1,448	QED	School
Ross et al. (1992)	1	2, 7	Local	Urban	Mixed	Daily	836	QED	School, social, Ext, Int, SP
Roukema (2005)	2	3	Local	Urban & Rural	Mixed	-	410	QED	School
Schinke, Cole, and Poulin (2000)	4	1, 2	National	Urban	Mixed	Daily	189	QED	School, Ext
Scurlock (2015)	1	1	Local	Urban	Non-academic	2 × week	89	QED	School, social, SP
Sibley-Butler (2004)	1	1, 2, 3	National	Rural	-	-	78	QED	School, Ext
Siddiqui et al. (2019)	1	1	National	-	Mixed	-	1,840	RCT	School, social
Smeallie (1997)	2	3	Local	-	Mixed	2 × week	62	RCT	School, SP
St. Pierre et al. (2001)	1	1, 2, 7	National	Urban	Mixed	Daily	126	RCT	School, social, Ext

TABLE 3 (Continued)

Study citation	Grade level ^a	Risk/marginalized characteristic ^b	Programme coverage	Programme location	Programme focus	Weekly contact	Total sample (N)	Study design ^c	Outcomes ^d
Tebes et al. (2007)	4	2	Local	Urban	Non-academic	-	304	QED	Ext
Updegraff et al. (2016)	1	1, 2	Local	-	Non-academic	1 × week	108	RCT	Social, Int
Venzen (2011)	2	1, 2, 3	National	Urban	Mixed	-	58	RCT	School
Weisman et al. (2003)	4	2	Local	-	Mixed	3–4 × week	1,070	QED	School, social, Ext
Welsh et al. (2002)	4	1, 2	National	Urban	-	Daily	65,779	QED	School

Note: N = 57 as Biggart et al. (2014) includes two reports (elementary and middle school data). -, Indicates information was missing from study.

^aGrade level: 1, elementary; 2, middle; 3, high; 4, mixed.

^bRisk/marginalized characteristic: 1, low income; 2, youth of colour; 3, low academic achievement/Title I school; 4, Single parent family; 5, Limited English proficiency; 6, experience trauma; 7, risky behaviours (truancy, substance use, delinquency).

^cRCT, Randomized control trial; QED, Quasi-experimental design.

^dOutcomes; Ext, externalizing; Int, internalizing; SP, self-perception.

TABLE 4 Moderator analyses of the effectiveness of ASPs

Moderator variable	k	#ES	B ₀ /g	SE ₀	B ₁	SE ₁	F (df ₁ , df ₂)
Outcome category							F (4,603) = 1.5786
Internalizing (RC)	9	19	0.1725	0.1659			
Externalizing	26	110	0.1450	0.0915	−0.0275	0.1695	
School-related	47	335	0.2313**	0.0710	0.0587	0.1660	
Self-perception/identity	17	65	0.2545*	0.1090	0.0819	0.1744	
Social	17	79	0.1544	0.1024	−0.0182	0.1696	
Youth characteristics							
Age	25	338	0.1695	0.5948	0.0094	0.0537	F (1,336) = 0.0309
Grade							F (3,611) = 5.0601
Elementary (RC)	22	194	0.1698	0.0995			
Middle	16	153	0.0588	0.1165	−0.1110	0.1532	
High school	2	41	0.1053	0.2966	−0.0645	0.3128	
Mixed grades	17	227	0.4131***	0.1155	0.2433	0.1524	
Race							
% White	35	419	0.3584**	0.1295	−0.4035	0.3303	F (1,417) = 1.4927
% Black	38	457	0.0091	0.1840	0.3950	0.2776	F (1,455) = 2.0236
% Latinx	32	392	0.3279*	0.1483	−0.3020	0.3480	F (1,390) = 0.7529
%Asian Amer./Pac. Island	30	354	0.3083*	0.1295	−0.6025	0.6905	F (1,352) = 0.7614
% indigenous	27	255	0.1841*	0.0847	−0.2522	0.4528	F (1,253) = 0.3103
% other race	30	294	0.1699*	0.0799	−0.3988	0.4901	F (1,292) = 0.6621
Gender (% male)	51	559	0.2095	0.2305	−0.0301	0.4273	F (1,557) = 0.005
% free/reduced lunch	28	240	0.1754	0.1222	0.0052	0.1510	F (1,238) = 0.0012
Program characteristics							
Program location							F (3,496) = 0.7104
Urban (RC)	41	451	0.2609**	0.0821			
Suburban	1	6	0.4058	0.5424	0.1449	0.5486	
Rural	4	33	0.1341	0.2666	−0.1268	0.2789	
Mixed	3	10	0.0271	0.3388	−0.2338	0.3486	
Program type							F (1,613) = 0.7141
National (RC)	17	212	0.2870*	0.1159			

TABLE 4 (Continued)

Moderator variable	k	#ES	B ₀ /g	SE ₀	B ₁	SE ₁	F (df ₁ , df ₂)
Local/regional	40	403	0.1700	0.0757	−0.1170	0.1384	F (2,577) = 1.6797
Program setting							
School (RC)	21	225	0.1254	0.1069			
Community-based	29	335	0.2807**	0.0932	0.1553	0.1419	F (2,567) = 0.1227
Mixed/unsure	3	20	0.0107	0.2842	−0.1147	0.3036	
Program focus							
Academic (RC)	11	92	0.1887	0.1515			F (4,546) = 3.1918
Non-academic	13	123	0.1824	0.1379	−0.0063	0.2049	
Mixed	29	355	0.2336*	0.0940	0.0449	0.1783	
Programme contact (session frequency)							
1×/week or less (RC)	5	58	0.2687	0.2178			F (3,611) = 1.2586
2×/week	9	105	0.1238	0.1716	−0.1449	0.2772	
3/4×/week	19	187	0.1093	0.1170	−0.1594	0.2472	
Daily	13	195	0.4174**	0.1456	0.1487	0.2620	F (1,613) = 0.2556
Other	1	6	0.3784	0.5151	0.1097	0.5592	
Implementation fidelity							
Programme operationalization							F (3,611) = 4.6308
Manualized (RC)	18	213	0.1203	0.1115			
General info	22	189	0.2546*	0.1032	0.1343	0.1519	
Very little description	16	190	0.2570*	0.1255	0.1367	0.1679	F (3,611) = 4.6308
No program description	1	23	−0.0246	0.4346	−0.1449	0.4487	
Fidelity measurement							
Not measured (RC)	43	266	0.1846*	0.0753			F (3,611) = 4.6308
Yes measured	14	349	0.2563*	0.1201	0.0716	0.1417	
Staff training							
No training (RC)	1	8	−0.2062	0.4582			F (3,611) = 4.6308
Some training	18	185	0.0659	0.1101	0.2721	0.4712	
Extensive training	10	107	0.1528	0.1444	0.3589	0.4804	
Unsure if training provide	28	315	0.3358***	0.0903	0.5419	0.4670	

(Continues)

TABLE 4 (Continued)

Moderator variable	k	#ES	B ₀ /g	SE ₀	B ₁	SE ₁	F (df ₁ , df ₂)
Methodological characteristics							
Publication type							
Published article (RC)	33	412	0.3145***	0.0798			F (2,612) = 4.7643
Dissertation	14	113	0.0238	0.1294	−0.2907	0.1521	
Report/other	10	90	0.0581	0.1531	−0.2564	0.1726	
Study design							
RCT (RC)	23	294	0.2013*	0.0970			F (1,613) = 0.0023
Quasi-experimental	34	321	0.2075*	0.0849	0.0062	0.1290	
Study sample size	57	615	0.2087**	0.0655	<0.001	<0.001	F (1,613) = 0.0608 F (2,612) = 3.3455
Comparison condition							
Nothing/waitlist (RC)	43	386	0.1591*	0.0723			
Treatment as usual	8	90	0.1981	0.1589	0.0390	0.1746	
Alternative program	6	139	0.5240**	0.1860	0.3649	0.1996	
Outcome measure type							
Archival/school record (RC)	39	222	−0.0060	0.0748			F (1,613) = 27.99***
Scale/survey/questionnaire	41	393	0.3597	0.0699	0.3657***	0.0691	
Outcome source							
Youth self-report (RC)	29	269	0.3654***	0.0771			F (5,609) = 31.36***
Parent-report	5	26	0.3760*	0.1714	0.0106	0.1727	
Teacher-report	13	76	0.4256***	0.1009	0.0602	0.0969	
Other person report	4	15	0.2123	0.2232	−0.1531	0.2219	
Official/school record	37	217	−0.0086	0.0765	−0.3740***	0.0784	
Other report	4	12	0.1152	0.2401	−0.2502	0.2446	

Abbreviations: #ES, number of effect sizes; B₀/g, intercept/mean effect size; B₁, estimated regression coefficient; F (df₁, df₂), omnibus test; k, number of studies; RC reference category; SE₀, standard error; SE₁, standard error.

p* < .05; *p* < .01; ****p* < .001.

moderators since these two variables were only cited in only seven studies (12.28% of total studies) and 10 studies (17.54% of total studies) respectively.

Results demonstrated statistically significant differences in the effect of ASPs based on outcome source ($F [5, 609] = 31.36, p < .001$), with official records ($B = -0.37, t = 0.08, p < .001$) yielding smaller effect sizes than self-report and teacher-reported outcomes. Outcome measure type was also a significant moderator of ASP effects ($F [1, 613] = 27.99, p < .001$), with rating scale/survey/checklist/questionnaire outcome measures yielding larger effect sizes ($B = 0.37, t = 0.07, p < .001$) than archival reports/school records. No significant moderating effects were found for any other methodological characteristics.

4.3 | Publication bias and sensitivity analyses

Across all included studies, there was generally high risk of bias for three of five indicators on the Cochrane risk-of-bias tool (selection bias, performance bias, detection bias; see Figure 2). Of 56 studies, selection bias (randomization) risk was rated as high in 35 (62.50%), low in 18 studies (32.14%), and unclear in three (5.36%). In 41 (73.21%) of the 56 included studies, performance bias risk was rated as high, 12 (21.43%) were rated as low, and 3 (5.36%) were rated as unclear. In terms of detection bias, 36 (64.29%) studies were rated as high, 10 (17.86%) were rated as low, and 10 (17.86%) as unclear. A complete summary of risk-of-bias coding across all included studies can be found in the electronic supplemental materials.

Furthermore, results of the funnel plot and trim-and-fill analyses suggest that publication bias was unlikely, but that selection bias was likely. There were no missing studies on the left side of the funnel (see Figure 3) and examining the distribution of effect sizes at the study level resulted in only a slightly smaller, significant overall mean effect size ($g = 0.1839, p < .001$). However, 18 studies were missing from the right side of the funnel plot, suggesting that there was possible selection bias in the included studies; analyses in the current study may have excluded studies with larger effect sizes, and as a result, the overall effect size of ASPs found in the current study may be an underestimation of their true effect. Accounting for this selection bias, a trim-and-fill analysis yielded a larger mean effect size of $g = 0.3187, p = <.001$.

Finally, findings from leave-one-out sensitivity analyses indicated that the overall effect size remained significant after each return, suggesting that none of the included studies had an individual, disproportionate impact on the overall findings. Moreover, the interval of effect sizes obtained through the sensitivity analyses ($0.1536 < g < 0.2170$) contains the overall effect size based on the total set of studies ($g = 0.2049$) and overlaps with the 95% confidence interval of the total effect size (95% CI: 0.0806–0.3293). Full results of leave-one-out sensitivity analyses can be found in the electronic supplementary materials.

5 | DISCUSSION

The goal of the current study was to investigate and synthesize the extant literature on the effects of ASPs on outcomes among youth with marginalized identities. This meta-analysis improves upon the limitations of prior meta-analyses by updating the existing literature and empirically assessing a more comprehensive range of outcomes and moderators with rigorous statistical methods among a specific population of youth.

Multi-level meta-analyses revealed an overall effect size of $g = 0.20$ across 56 studies, 615 effect sizes, and 128,538 youth participants. This finding represents a significant but small effect according to Cohen's (1998) standards and is consistent with previous studies of the effects of ASPs for youth of all identities (e.g., Ciocanel et al., 2017; Durlak et al., 2010; Lauer et al., 2006). Simply put, these results suggest that ASPs do have the potential to bring about some positive change across a range of developmental domains including social-emotional/interpersonal skills, mental and behavioural health, school success, and identity development for youth who engage in these

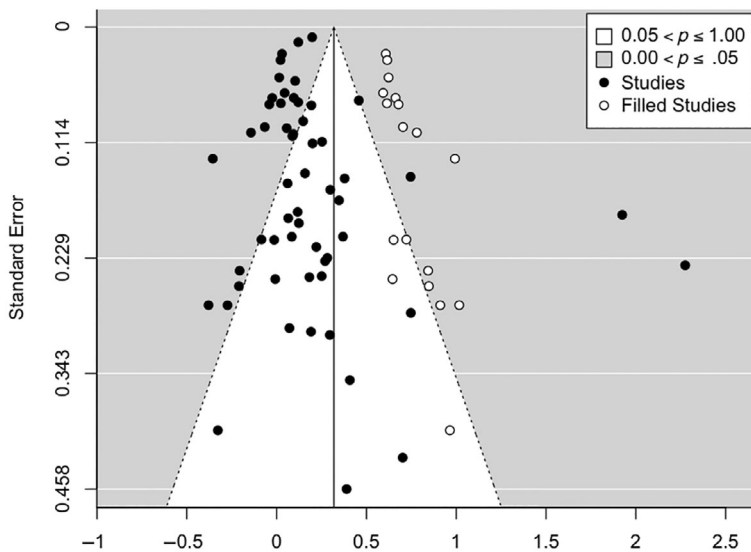


FIGURE 3 Funnel plot after trim-and-fill analysis. No studies were missing on the left side of the funnel plot, indicating minimal risk of publication bias. Eighteen studies were missing on the right side of the funnel, indicating possible selection bias.

types of programmes. Of note, specific youth outcome category did not significantly moderate overall effects. This finding could indicate that ASPs may not be particularly effective at improving any one targeted youth outcome, but have some promise for youths' overall general improvement. Alternatively, it is possible that primary evaluations of ASPs are unable to accurately capture the full nuance of counterfactual conditions (i.e., youth who are not engaged in ASPs or other community programmes). Often, the ASP under evaluation is assessed in a vacuum, with outside activities and supports of youth in both the intervention and control groups unmeasured. It is possible that youth in both the treatment and control conditions are also engaging with external community programmes and supports that are unaccounted for. Thus, a representative counterfactual is needed to accurately capture the full nuance of conditions where ASPs are not implemented in order to understand their effectiveness in the opposite direction (i.e., not just how many positive outcomes were or were not achieved, but how many negative outcomes were avoided). Related to this, it may also be useful for future research to further assess and compare ASPs operating at different levels of the prevention continuum. For example, future studies may seek to investigate whether ASPs functioning as universal prevention programmes are effective in delaying or reducing negative trends in youth functioning and outcomes. Alternatively, if the goal of ASPs is to improve specific outcomes for youth, it will be important for future practice and research to become more comprehensive in their assessment of youth experiences and characteristics and more targeted in their evaluation of programme practices.

Notably, this effect size represents a more impressive improvement within the broader context of typical educational interventions. According to the (What Works Clearinghouse, 2014) 'improvement index,' interpreted as the practical importance of a programme's effectiveness, this effect size represents an improvement of a little less than nine percentile points (2014), which is indeed not negligible. The improvement index reflects 'the expected change in percentile rank for an average comparison group student if the student had received the intervention' (p. 21). Thus, when understanding effect sizes through metrics other than Cohen's *d*, this relatively modest effect may hold more practical significance, speaking to the potential importance of ASPs as positive and effective developmental spaces in youth's lives.

Results of moderator analyses revealed that outcome source and outcome measure type were significant moderators of the effectiveness of ASPs. Outcome measures reported by youth (i.e., self-report) or teachers were more likely to yield larger effect sizes, whereas official records yielded smaller effect sizes. For example, one included study contained youth-report measures, teacher-report measures, as well as official records. Youth- and teacher-report measures focused more on degree of youth engagement and enjoyment in learning about various school subjects, whereas official reports documented youths' actual grades for specific subjects (Schinke et al., 2000). Likewise, archival reports and school records yielded smaller effect sizes than those derived from other measure types (i.e., rating scales, surveys, checklists, questionnaires). For example, some studies used rating scales or surveys to assess constructs such as youth motivation, perceived competence, and goals (Grolnick et al., 2007). Naturally, this finding may reflect the phenomenon of self-report bias in research (Paulhus & Vazire, 2007) or a mismatch between youths' perceived performance and metrics like report card grades or grade point average. Moreover, this result is consistent with findings from studies showing that methodological features affect effect sizes. Specifically, researcher-developed tests tend to yield higher effect sizes than standardized tests and instruments (Cheung & Slavin, 2016). It is also possible that outcomes represented in more narrow, quantifiable archival or school reports (i.e., grades and attendance) do not reflect the different or broader nuanced gains that youth may be making in other outcomes that are typically assessed by self-report scales, checklists, and surveys (such as self-perceptions about competence and motivation). Indeed, in one meta-review of 17 What Works Clearinghouse educational intervention review papers, Slavin and Madden (2011) found larger effect sizes for outcome measures that were more aligned with or inherent to the actual intervention than measures that were less related to the specific intervention being assessed. For example, measures inherent to the experimental interventions included specific content or skills that only the experimental group (not the control/comparison group) had access to. Measures classified as not inherent to the intervention included standardized or otherwise specialized measures that were not developed by the study authors/researchers, or content that was presented to both experimental and control groups. The same notion may hold true for findings in this study, depending on whether the youth outcomes assessed were actually aligned with the interventions and activities in the ASPs.

Given the significant, substantial heterogeneity detected in this study, it may be surprising that the results revealed no other significant moderating variables. However, this finding is highly consistent with previously published meta-analyses on ASPs, which have also failed to identify significant moderators of after-school effects despite high heterogeneity and many variables being examined. The lack of moderators reflects the phenomenon of the 'black box' of ASPs that has been documented in other studies (e.g., Yohalem & Wilson-Ahlstrom, 2010), and the lack of specificity behind how these programmes produce positive outcomes.

A significant issue that remains in the literature base on the effects of ASPs on youth outcomes is that studies often leave out critical information that could be used to conduct additional moderator analyses. Most notable is the lack of information about programme staff, despite staff being widely accepted as a key mechanism through which youth develop positive outcomes in ASPs (Rhodes, 2004). Information related to staff demographics, education, training, and supervision are all necessary variables to explore to better understand and contextualize both the experiences and outcomes resulting from youth engagement in ASPs. For example, research suggests that staff identity may be important for youth, particularly in programmes comprised primarily of youth of colour. Shared identity or community may increase trust between staff and youth when building relationships (Griffith, 2014; Simpkins et al., 2017). Ample research also suggests the important role of staff experience, education, training, and quality for youths' experiences in ASPs and thus their effect on youth outcomes (Astroth et al., 2004). In addition to many evaluations not reporting these staff characteristics, programme activities and information about the operationalization of the programme (and as a result, implementation adherence, fidelity, and quality) are unspecified. Thus, although we can conclude that participation in ASPs results in a small but significant overall positive effect on youth outcomes over being in no organized programming, it was not possible to account for the exact nature of the evaluated programmes.

Additionally, the moderating role of other supportive interventions (e.g., tutoring, psychotherapy, formal/natural mentors, neighbourhood support, other after-school activities) has remained unaccounted for in the majority of ASP evaluations and meta-analyses. Assessing the impact of youths' networks of support and other programming (especially for youth in comparison conditions) will advance our understanding of the contribution of ASPs in youth outcomes. Future studies should ensure that comprehensive information is reported on programme, youth, and staff characteristics. Given the heterogeneity of ASPs, it will remain challenging to draw generalizable conclusions about their effectiveness, particularly in the absence of systematic specification and reporting of important intervention components. To account for heterogeneity across youth programmes, McQuillin et al. (2020) have argued for evaluating the effect sizes of specific programme practices as opposed to overall programme effects.

Finally, it is possible that the non-significant moderators simply do not play a major role in influencing the magnitude of the effects of ASPs. However, given the heterogeneity in ASPs, other moderators that have previously been unable to be coded may act as proxy variables and deserve further investigation. For example, programme location (i.e., urban, suburban, rural) and type (i.e., national, local) could serve as a potential proxy for programme funding per student, which is typically unreported in empirical studies. Likewise, the type of programme (i.e., academic, recreational, mixed activities) may be less consequential than whether programme activities are targeted to the needs of youth entering the programme (Christensen et al., 2020).

5.1 | Study limitations and strengths

There are limitations to the current study that should be acknowledged when interpreting the results. First, analyses were limited by the information reported in the primary studies included in this meta-analysis as a result of primary studies lacking information on variables such as staff level of education, measurement of implementation fidelity, and youth's actual programme attendance. It will be necessary for future ASP evaluations to include detailed information about key factors that may moderate overall effect sizes. Recommendations include reporting information about ASP staff training and characteristics, detailed information on programme activities, implementation fidelity, and youth attendance, as well as information about constellations of care or support youth may be engaging within their communities outside of ASPs.

Furthermore, although no publication bias was detected, likely resulting from the inclusion of both published and unpublished works, there was evidence of selection bias, and a moderate portion of included studies posed a risk of bias. For example, the inclusion of quasi-experimental design studies helped increase sample size, statistical power, and programme representativeness, but may have allowed for bias related to youths' self-selection into ASPs, such that more extraverted or well-supported youth may have disproportionately participated and positively influenced overall effects. Despite this potential, the current meta-analysis also included randomized-controlled trial designs, thus reducing the likelihood of self-selection bias. Regardless, after conducting a recent meta-analysis of primarily unpublished reports with high risk of bias, Lester et al. (2020) noted the importance of examining less rigorous studies, as they are more reflective of the current literature base on ASPs and how programmes operate in the 'real world.' It will be important to balance the notion of ecological validity against the need for more rigorous and specific programme evaluation.

Despite these limitations, this study had a number of strengths. The use of multi-level meta-analysis represents a highly robust meta-analytic method, allowing for variance to be accounted for from three levels (as opposed to typically one source). Each stage of this systematic review and meta-analysis also adhered to quality methods outlined by Cochrane and the Campbell Collaboration, representing significant methodological improvements from previous meta-analyses on the effects of ASPs. Additionally, analyses in the current study examined youth who identified with characteristics associated with social marginalization (e.g., race/ethnicity, low academic performance, etc.). Although this may somewhat limit the generalizability of the results, the intentional selection of this sample provided a specific and comprehensive examination of a particularly salient group of youth often targeted for involvement in ASPs.

Finally, a comprehensive range of youth outcomes and moderating variables was examined among the most up-to-date set of studies, allowing for a thorough assessment of the effects of ASPs and potential factors influencing their magnitude. These outcomes and moderators were also examined specifically among a population of marginalized youth who, despite being a historically targeted group for after-school engagement, are an understudied population in existing meta-analytic studies.

6 | IMPLICATIONS AND RECOMMENDATIONS FOR RESEARCH, PRACTICE, AND POLICY

Findings of this study point to the potential importance of ASPs, but also the need for greater specificity in the intentions, goals, and evaluations of ASPs, including the inputs and moderating processes that are thought to be associated with specific outcomes across primary after-school evaluation studies. In the current study, over 80% of included studies did not specify key inputs such as staff education and details about fidelity measurement. Thus, findings reveal one overall effect size for ASPs, but offer little direction on what levers to pull to improve outcomes. Greater specification will enable programmes to identify, implement, adhere to, and evaluate the inputs and processes that are put into place to achieve desired goals.

It is possible that the lack of clarity around inputs, moderators, and outcomes reflects a broader lack of clarity surrounding the function and goals of ASPs, leading many programmes to cast a wide net for unspecified outcomes (McQuillin et al., 2020). Indeed, there remains considerable debate regarding the extent to which ASPs should be designed to compensate for particular challenges (e.g., insufficient school-work/homework time, emotional and behavioural challenges) versus providing opportunities for positive youth development (e.g., relationship formation, identity exploration, interest discovery). The indicators and relative effectiveness of these different approaches to service delivery deserve careful consideration as they have important implications for how programmes are structured and evaluated.

In reflecting on the need for increased clarity around programme intentions and goals, two possible directions for the future of ASPs emerge: (1) one in which positive youth development (PYD) is balanced with prevention science such that ASPs become centres for the application of evidence-based skills and targeted care for youth or, (2) one in which ASPs are viewed as one developmental context within a constellation of support, in which the primary goal is for youth to have fun, develop identity and critical consciousness, and build networks of social support. Both of these possibilities include a focus on individual youth strengths and societal-level structural oppression. These two directions also need not be mutually exclusive. For youth experiencing more challenges, targeted care in ASP settings may be useful. For youth who are not experiencing as many struggles but would benefit from more caring adults (i.e., natural mentoring relationships) and opportunities to explore identity, the second direction may help fill those gaps. Moreover, these two directions could, indeed, co-exist in which primarily recreational ASPs provide services such as embedded tutoring or paraprofessional mentors and mental health care on certain days. In both of these possible directions, the key will involve being mindful of what inputs and processes are optimal for specific youth outcomes.

Given the heightened rates of structural oppression (e.g., poverty, racism) as well as mental health, academic, and behavioural challenges among youth, recent research in the adjacent field of youth mentoring has been considering similar models in which mentoring programmes are re-imagined as contexts for paraprofessionals (i.e., volunteer mentors) to apply evidence-based care for youth (e.g., Christensen et al., 2020; Hart, Sung, McQuillin, & Schleider, 2021; Rhodes, 2020). It is noteworthy that ASPs yield almost identical overall effect sizes as mentoring programmes, but they serve more youth in a less expensive and more accessible way. Thus, ASPs may represent a more scalable and equally effective approach to improving youth outcomes over one-on-one delivery of PYD activities in the mentoring context. ASPs receive more federal funding (Rhodes, 2020) and most mentored youth are already involved in PYD programmes (Jarjoura, Tanyu, Forbush, Herrera, & Keller, 2018). Thus, if the goal

is to create more youth-adult bonds and PYD activities, the most efficient structure would involve practice and policy changes that lead to more caring adults in ASPs to improve youth-staff ratios, staff burnout and turnover, and ideally, support longer tenure of staff through better pay.

7 | CONCLUSION

The current study aimed to address multiple existing gaps in the literature through a meta-analysis of the overall effects of ASPs on internalizing (i.e., mental health), externalizing (i.e., behavioural), school-related, social functioning and self-perception/identity outcomes among youth with marginalized identities. Findings suggest that ASPs have a small, yet significant positive effect on youths' outcomes. Future research on ASPs will be strengthened by increasing the transparency of information reported in programme evaluation studies to better understand the role of moderating variables of ASP effectiveness. Implications for practice include clarifying the intentions, goals, and processes of ASPs to better assess their potential benefits for youth.

CONFLICT OF INTEREST

The authors have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- Afterschool Alliance. (2014a). America after 3PM: Afterschool programs in demand. http://afterschoolalliance.org/documents/AA3PM2014/AA3PM_National_Report.pdf
- Afterschool Alliance. (2014b). Keeping kids safe and supported in the hours after school. http://afterschoolalliance.org/documents/issue_briefs/issue_KeepingKidsSafe_65.pdf
- Arcaira, E., Vile, J. D., & Reisner, E. R. (2010). *Achieving high school graduation: Citizen Schools' youth outcomes in Boston*. Washington, DC: Policy Studies Associates Inc.
- Assink, M., & Wibbelink, C. J. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, 12(3), 154–174.
- Astroth, K. A., Garza, P., & Taylor, B. (2004). Getting down to business: Defining competencies for entry-level youth workers. *New Directions for Youth Development*, 2004(104), 25–37.
- Averett, P., Crowe, A., & Hall, C. (2015). The youth public arts program: Interpersonal and intrapersonal outcomes for at-risk youth. *Journal of Creativity in Mental Health*, 10(3), 306–323.
- Avery, C. (2013). *Evaluation of the college possible program: Results from a randomized controlled trial* (No. w19562). National Bureau of Economic Research.
- Baker, D., & Witt, P. A. (1996). Evaluation of the impact of two after-school programs. *Journal of Park and Recreation Administration*, 14(3), 60–81.
- Belgrave, F. Z., Chase-Vaughn, G., Gray, F., Addison, J. D., & Cherry, V. R. (2000). The effectiveness of a culture and gender-specific intervention for increasing resiliency among African American preadolescent females. *Journal of Black Psychology*, 26(2), 133–147.
- Bergin, D. A., Hudson, L. M., Chryst, C. F., & Resetar, M. (1992). An afterschool intervention program for educationally disadvantaged young children. *The Urban Review*, 24(3), 203–217.
- Bialeschki, M. D., Henderson, K. A., & James, P. A. (2007). Camp experiences and developmental outcomes for youth. *Child and Adolescent Psychiatric Clinics of North America*, 16(4), 769–788.
- Biggart, A., Kerr, K., O'Hare, L., & Connolly, P. (2013). A randomised control trial evaluation of a literacy after-school programme for struggling beginning readers. *International Journal of Educational Research*, 62, 129–140.
- Biggart, A., Sloan, S., & O'Hare, L. (2014). *A Longitudinal Follow Up Study of the Doodle Den After School Programme*. Dublin: Childhood Development Initiative (CDI).
- Blumer, M. L. C., & Werner-Wilson, R. J. (2010). Leaving no girl behind: Clinical intervention effects on adolescent female academic "high-risk" behaviors. *Journal of Feminist Family Therapy*, 22(1), 22–42.

- Bowman, B. T., Donovan, S. M., & Burns, M. S. (2000). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.
- Brock, L. L., Murrah, W. M., Cottone, E. A., Mashburn, A. J., & Grissmer, D. W. (2018). An after-school intervention targeting executive function and visuospatial skills also improves classroom behavior. *International Journal of Behavioral Development*, 42(5), 474–484.
- Burton, L. M., Mattingly, M., Pedroza, J., & Welsh, W. (2017). State of the union 2017: Poverty. Stanford Center on Poverty and Inequality. https://inequality.stanford.edu/sites/default/files/Pathways_SOTU_2017_poverty.pdf
- Campbell Collaboration. (2014). Campbell Collaboration systematic reviews: Policies and guidelines. http://www.campbellcollaboration.org/lib/download/3308/C2_Policies_Guidelines_Version_1_0.pdf
- Carr, E. M. (2015). Afterschool program interventions that support the academic achievement, behavior, and engagement of at-risk student populations [Doctoral dissertation, University of Houston Clear Lake].
- Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292.
- Christensen, K. M., Hagler, M. A., Stams, G. J., Raposa, E. B., Burton, S., & Rhodes, J. E. (2020). Non-specific versus targeted approaches to youth mentoring: A follow-up meta-analysis. *Journal of Youth and Adolescence*, 49(5), 959–972.
- Ciocanel, O., Power, K., Eriksen, A., & Gillings, K. (2017). Effectiveness of positive youth development interventions: A meta-analysis of randomized controlled trials. *Journal of Youth and Adolescence*, 46(3), 483–504.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Crombie, I. K., & Davies, H. T. (2009). What is meta-analysis. *What Is*, 1–8. <https://www.betterevaluation.org/sites/default/files/Meta-An.pdf>
- Developmental Services Group, Inc. (2010). *Afterschool programs*. Literature review. Washington, D.C.: Office of Juvenile Justice and Delinquency Prevention. https://www.ojjdp.gov/mpg/litreviews/Afterschool_Programs.pdf
- Diaz, J. D. (2005). School attachment among rural Latino youth in Minnesota. *Hispanic Journal of Behavioral Sciences*, 27, 300–318.
- Dotterer, A. M., McHale, S. M., & Crouter, A. C. (2007). Implications of out-of-school activities for school engagement in African American adolescents. *Journal of Youth and Adolescence*, 36(4), 391–401.
- Dreyer, K. J. (2010). An examination of academic outcomes for students who attend a school-based afterschool program [Doctoral dissertation, University of Pittsburgh].
- Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, 45(3–4), 294–309.
- Durlak, J. A., & Weissberg, R. P. (2007). The impact of after school programs that promote personal and social skills. Collaborative for Academic, Social, and Emotional Learning, University of Illinois at Chicago. http://www.casel.org/downloads/ASP_full.pdf
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot based method of using and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463.
- Eccles, J. S., Barber, B. L., Stone, M., & Hunt, J. (2003). Extracurricular activities and adolescent development. *Journal of Social Issues*, 59(4), 865–889.
- EdBuild. (2019). \$23 billion. <https://edbuild.org/content/23-billion/full-report.pdf>
- Foley, E. M., & Eddins, G. (2001). *Preliminary analysis of Virtual Y after-school program participants' patterns of school attendance and academic performance*. New York, NY: Fordham University.
- Forrest-Bank, S. S., Nicotera, N., Bassett, D. M., & Ferrarone, P. (2016). Effects of an expressive art intervention with urban youth in low-income neighborhoods. *Child and Adolescent Social Work Journal*, 33(5), 429–441.
- Frazier, S., Mehta, T., Atkins, M., Hur, K., & Rusch, D. (2013). Not just a walk in the park: Efficacy to effectiveness for after school programs in communities of concentrated urban poverty. *Administration and Policy in Mental Health and Mental Health Services Research*, 40(5), 406–418.
- Frazier, S. L., Cappella, E., & Atkins, M. S. (2007). Linking mental health and after school systems for children in urban poverty: Preventing problems, promoting possibilities. *Administration and Policy in Mental Health and Mental Health Services Research*, 34(4), 389–399. <https://doi.org/10.1007/s10488-007-0118-y>
- Gottfredson, D., Cross, A., Wilson, D., Rorie, M., & Connell, N. (2010). An experimental evaluation of the all stars prevention curriculum in a community after school setting. *Prevention Science*, 11(2), 142–154.
- Gottfredson, D., Gerstenblith, S. A., Soulé, D., Womer, S. C., & Lu, S. (2004). Do after school programs reduce delinquency? *Prevention Science*, 5(4), 253–266.
- Graham, S., Taylor, A., & Hudley, C. (2015). A motivational intervention for African American boys labeled as aggressive. *Urban Education*, 50(2), 194–224.
- Grassi, A. M. (2012). Quality After-School Programming and its Relationship to Achievement Related Behaviors and Academic Performance [Doctoral dissertation, Cleveland State University].

- Green, H. K. (2010). The impact of an academic sports-mentoring afterschool program on academic outcomes in at-risk youth [Doctoral thesis, Drexel University].
- Griffith, A. (2014). The formation of adolescents' trust in adult leaders at project-based youth programs (Doctoral dissertation, University of Illinois at Urbana-Champaign).
- Grolnick, W. S., Farkas, M. S., Sohmer, R., Michaels, S., & Valsiner, J. (2007). Facilitating motivation in young adolescents: Effects of an after-school program. *Journal of Applied Developmental Psychology*, 28(4), 332–344.
- Halpern, R. (2005). Instrumental relationships: A potential relational model for inner-city youth programs. *Journal of Community Psychology*, 33(1), 11–20.
- Hart, M. J., Sung, J., McQuillin, S. D., & Schleider, J. L. (2021). Expanding the reach of psychosocial services for youth: Untapped potential of mentor-delivered single session interventions.
- Helseeth, S. A., & Frazier, S. L. (2018). Peer-assisted social learning for diverse and low-income youth: Infusing mental health promotion into urban after-school programs. *Administration and Policy in Mental Health and Mental Health Services Research*, 45(2), 286–301.
- Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558.
- Higgins, J. P., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Vol. 4). John Wiley & Sons. The Cochrane Collaboration. https://handbook-5-1.cochrane.org/chapter_9/9_7_sensitivity_analyses.htm
- Hirsch, B., Hedges, L. V., Stawicki, J., & Mekinda, M. A. (2011). *After-school programs for high school students: An evaluation of After School Matters*. Chicago, IL: Northwestern University.
- Hirsch, B. J. (2005). *A place to call home: After-school programs for urban youth*. New York, NY: Teachers College Press.
- Hirschman, C., & Voloshin, I. (2007). The structure of teenage employment: Social background and the jobs held by high school seniors. *Research in Social Stratification and Mobility*, 25(3), 189–203.
- Hofferth, S. L., & Jankuniene, Z. (2001). Life after school. *Educational Leadership*, April, 19–23.
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, 11(2), 193–206.
- Hynes, K., & Sanders, F. (2011). Diverging experiences during out-of-school time: The race gap in exposure to after-school programs. *The Journal of Negro Education*, 80(4), 464–476.
- James-Burdumy, S., Dynarski, M., & Deke, J. (2008). After-school program effects on behavior: Results from the 21st Century Community Learning Centers program national evaluation. *Economic Inquiry*, 46(1), 13–18.
- James-Burdumy, S., Dynarski, M., & Deke, J. (2007). When elementary schools stay open late: Results from the national evaluation of the 21st Century Community Learning Centers program. *Educational Evaluation and Policy Analysis*, 29(4), 296–318.
- Jarjoura, G. R., Tanyu, M., Forbush, J., Herrera, C., & Keller, T. E. (2018). Evaluation of the mentoring enhancement demonstration program: Technical report. National Criminal Justice Reference Center.
- Jenner, E., & Jenner, L. W. (2007). Results from a first-year evaluation of academic impacts of an after-school program for at-risk students. *Journal of Education for Students Placed at Risk*, 12(2), 213–237.
- Kaufman, C. E., Schwinn, T. M., Black, K., Keane, E. M., Big Crow, C. K., Shangreau, C., ... Morse, B. (2018). Impacting precursors to sexual behavior among young American Indian adolescents of the Northern Plains: a cluster randomized controlled trial. *The Journal of Early Adolescence*, 38(7), 988–1007.
- Kremer, K. P., Maynard, B. R., Polanin, J. R., Vaughn, M. G., & Sarteschi, C. M. (2015). Effects of after-school programs with at-risk youth on attendance and externalizing behaviors: A systematic review and meta-analysis. *Journal of Youth and Adolescence*, 44(3), 616–636.
- La Torre, D., Leon, S., Wang, J., & Cai, L. (2019). Long-Term Outcome Study: A Longitudinal Study of LA's BEST Students' Persistence and Graduation Rates. CRESST Report 862. National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- LaFrance, S., Twersky, F., Latham, N., Foley, E., Bott, C., & Lee, L. (2001). *A safe place for healthy youth development: A comprehensive evaluation of the Bayview Safe Haven*. San Francisco, CA: BTW Consultants & LaFrance Associates.
- Langberg, J., Smith, B. H., Bogle, K. E., Schmidt, J. D., Cole, W. R., & Pender, C. A. S. (2007). A pilot evaluation of small group challenging horizons program (CHP): A randomized trial. *Journal of Applied School Psychology*, 23(1), 31–58.
- Langberg, J. M., Epstein, J. N., Urbanowicz, C. M., Simon, J. O., & Graham, A. J. (2008). Efficacy of an organization skills intervention to improve the academic functioning of students with attention-deficit/hyperactivity disorder. *School Psychology Quarterly*, 23(3), 407.
- Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, 76(2), 275–313.
- Le, T., Arifuku, I., Vuong, L., Tran, G., Lustig, D., & Zimring, F. (2011). Community mobilization and community-based participatory research to prevent youth violence among Asian and immigrant populations. *American Journal of Community Psychology*, 48(1/2), 77–88.

- LeCroy, C. W., McCullough Cosgrove, J., Cotter, K., & Fordney, M. (2018). Go Grrrls: A randomized controlled trial of a gender-specific intervention to reduce sexual risk factors in middle school females. *Health Education & Behavior, 45*(2), 286–294.
- Lester, A. M., Chow, J. C., & Melton, T. N. (2020). Quality is critical for meaningful synthesis of afterschool program effects: A systematic review and meta-analysis. *Journal of Youth and Adolescence, 49*(2), 369–382.
- Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin, 126*(2), 309–337.
- Mahoney, J. L., & Stattin, H. (2000). Leisure activities and adolescent antisocial behavior: The role of structure and social context. *Journal of Adolescence, 23*(2), 113–127.
- Marsh, H., & Kleitman, S. (2002). Extracurricular school activities: The good, the bad, and the nonlinear. *Harvard Educational Review, 72*(4), 464–515.
- Marshall, N. L., Coll, C. G., Marx, F., McCartney, K., Keefe, N., & Ruh, J. (1997). After- school time and children's behavioral adjustment. *Merrill-Palmer Quarterly (1982-), 49*(7), 497–514.
- Maynard, B. R., Peters, K. E., Vaughn, M. G., & Sarteschi, C. M. (2013). Fidelity in after- school program intervention research: A systematic review. *Research on Social Work Practice, 23*(6), 613–623.
- McComb, E. M., & Scott-Little, C. (2003). *After-school programs: Evaluations and outcomes*. Greensboro, NC: SERVE.
- McKinney, A. D. (1995). The effects of an after-school tutorial and enrichment program on the academic achievement and self-concept of below grade level first and second grade students [Doctoral dissertation, University of Mississippi].
- McQuillin, S. D., Lyons, M. D., Clayton, R. J., & Anderson, J. R. (2020). Assessing the impact of school-based mentoring: Common problems and solutions associated with evaluating nonprescriptive youth development programs. *Applied Developmental Science, 24*(3), 215–229.
- Miller, B. M. (2003). *Critical hours: Afterschool programs and educational success*. Quincy, MA: Nellie Mae Education Foundation. <https://www.nmefoundation.org/getmedia/08b6e87b-69ff-4865-b44e-ad42f2596381/Critical-Hours>
- Moldow, E. (2007). After-school program activities and academic achievement: A study in one urban k-8 school. Brandeis University, The Heller School for Social Policy and Management.
- Molina, B. S., Flory, K., Bukstein, O. G., Greiner, A. R., Baker, J. L., Krug, V., et al. (2008). Feasibility and preliminary efficacy of an after-school program for middle schoolers with ADHD: A randomized trial in a large public middle school. *Journal of Attention Disorders, 12*(3), 207–217.
- Nance, E. E., Moore, D. H., & Lewis, C. (2000). 21st century community learning centers: Do they affect student achievement? *Community Education Journal, 27*, 7–11.
- Nguyen, D. (2007). A statewide impact study of 21st Century Community Learning Center programs in Florida. [Doctoral dissertation, Florida State University].
- O'Hare, L., Biggart, A., Kerr, K., & Connolly, P. (2015). A randomized controlled trial evaluation of an after-school prosocial behavior program in an area of socioeconomic disadvantage. *The Elementary School Journal, 116*(1), 1–29.
- Office of Adolescent Health. (2019). Adolescent Mental Health Basics. https://www.hhs.gov/ash/oah/adolescent-development/mental-health/adolescent-mental-health-basics/index.html#_ftn1
- Oyserman, D., Terry, K., & Bybee, D. (2002). A possible selves intervention to enhance school involvement. *Journal of Adolescence, 25*(3), 313.
- Pastchal-Temple, A. (2012). The effect of regular participation in an after-school program on student achievement, attendance, and behavior. [Doctoral dissertation, Mississippi State University].
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). New York City, NY: Guilford Press.
- Pedersen, S., & Seidman, E. (2005). Contexts and correlates of out-of-school activity participation among low-income urban adolescents. In J. L. Mahoney, R. W. Larson, & J. S. Eccles (Eds.), *Organized activities as contexts of development* (pp. 85–109). Mahwah, NJ: Erlbaum.
- Pettway, D. D. (2005). An evaluation of SmartLinks (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3173506).
- Phillips, R. S. (1999). Intervention with siblings of children with developmental disabilities from economically disadvantaged families. *Families in Society, 80*(6), 569–577.
- Pierce, L. H., & Shields, N. (1998). The Be A Star community-based after-school program: Developing resiliency factors in high-risk preadolescent youth. *Journal of Community Psychology, 26*(2), 175–183.
- Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. *Educational Researcher, 42*(8), 424–434.
- Prenovost, J. K. E. (2001). A first-year evaluation of after-school learning programs in four urban middle schools in the Santa Ana Unified School District. [Doctoral dissertation, University of California—Irvine].
- Raposa, E. B., Rhodes, J., Stams, G. J. J., Card, N., Burton, S., Schwartz, S., ... Hussain, S. (2019). The effects of youth mentoring programs: A meta-analysis of outcome studies. *Journal of Youth and Adolescence, 48*(3), 423–443.

- Rhodes, A. M., & Schechter, R. (2014). Fostering resilience among youth in inner city community arts centers: The case of the artists collective. *Education and Urban Society*, 46(7), 826–848.
- Rhodes, J. E. (2004). The critical ingredient: Caring youth-staff relationships in after-school settings. *New Directions for Youth Development*, 2004(101), 145–161.
- Rhodes, J. E. (2020). *Older and Wiser*. Cambridge, MA: Harvard University Press.
- Riggs, N. R., Bohnert, A. M., Guzman, M. D., & Davidson, D. (2010). Examining the potential of community-based after-school programs for Latino youth. *American Journal of Community Psychology*, 45(3–4), 417–429.
- Riggs, N. R., & Greenberg, M. T. (2004). After-school youth development programs: A developmental-ecological model of current research. *Clinical Child and Family Psychology Review*, 7(3), 177–190.
- Ross, J. G., Saavedra, P. J., Shur, G. H., Winters, F., & Felner, R. D. (1992). The effectiveness of an after-school program for primary grade latchkey students on precursors of substance abuse. *Journal of Community Psychology*, 20(1), 22–38.
- Roth, J. L., Malone, L. M., & Brooks-Gunn, J. (2010). Does the amount of participation in afterschool programs relate to developmental outcomes? A review of the literature. *American Journal of Community Psychology*, 45(3–4), 310–324.
- Roukema, R. A. (2005). The impact of the Support Our Students (SOS) after-school program on the achievement of middle-grade students at risk of academic failure. [Doctoral dissertation, North Carolina State University].
- Schinke, S., Cole, K. C., & Poulin, S. R. (2000). Enhancing the educational achievement of at-risk youth. *Prevention Science*, 1(1), 51–60.
- Scott-Little, C., Hamann, M. S., & Jurs, S. G. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, 23(4), 387–419.
- Scurlock, D. (2015). Effect of occupation-based groups on self-concept of children aged 5–8: A pilot study. *Occupational Therapy in Health Care*, 29(1), 63–76.
- Sheroff, D. J., & Vandell, D. L. (2007). Engagement in after-school program activities: Quality of experience from the perspective of participants. *Journal of Youth and Adolescence*, 36(7), 891–903.
- Sibley-Butler, O. (2004). The impact of a 21st Century Community Learning Center after-school program on school-age children in grades 4 & 5. [Doctoral dissertation, University of Southern Mississippi].
- Siddiqui, N., Gorard, S., & See, B. H. (2019). Can learning beyond the classroom impact on social responsibility and academic attainment? An evaluation of the Children's University youth social action programme. *Studies in Educational Evaluation*, 61, 74–82.
- Simpkins, S. D., Riggs, N. R., Ngo, B., Vest Ettekal, A., & Okamoto, D. (2017). Designing culturally responsive organized after-school activities. *Journal of Adolescent Research*, 32(1), 11–36.
- Slavin, R. E., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness*, 4, 370–380.
- Smeallie, J. E. (1997). An evaluation of an after-school tutorial and study skills program for middle school students at risk of academic failure. [Doctoral Dissertation, University of Maryland at College Park].
- Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment fidelity in applied educational research: Expanding the adoption and application of measures to ensure evidence-based practice. *Education and Treatment of Children*, 30, 121–134. <https://doi.org/10.1353/etc.2007.0033>
- St. Pierre, T. L., Mark, M. M., Kaltreider, D. L., & Campbell, B. (2001). Boys & girls clubs and school collaborations: A longitudinal study of a multicomponent substance abuse prevention program for high-risk elementary school children. *Journal of Community Psychology*, 29(2), 87–106.
- Taheri, S. A., & Welsh, B. C. (2016). After-school programs for delinquency prevention: A systematic review and meta-analysis. *Youth Violence and Juvenile Justice*, 14(3), 272–290.
- Tebes, J. K., Feinn, R., Vanderploeg, J. J., Chinman, M. J., Shepard, J., Brabham, T., ... Connell, C. (2007). Impact of a positive youth development program in urban after-school settings on the prevention of adolescent substance use. *Journal of Adolescent Health*, 41(3), 239–247.
- U.S. Department of Education. (2014). Programs: 21st Century Community Learning Centers. <http://www2.ed.gov/programs/21stccclc/index.html>
- Updegraff, K. A., Umaña-Taylor, A. J., Rodríguez De Jesús, S. A., McHale, S. M., Feinberg, M. F., & Kuo, S. I. (2016). Family-focused prevention with Latinos: What about sisters and brothers? *Journal of Family Psychology*, 30(5), 633.
- Van Voorhees, B. W., Paunesku, D., Kuwabara, S. A., Basu, A., Gollan, J., & Hankin, B. L. (2008). Protective and vulnerability factors predicting new-onset depressive episode in a representative of U.S. adolescents. *Journal of Adolescent Health*, 42(6), 605–616. <https://doi.org/10.1016/j.jadohealth.2007.11.135>
- Vandell, D. L., & Posner, J. (1999). Conceptualization and measurement of children's after-school environments. In S. L. Freeman & T. D. Wachs (Eds.), *Assessment of the environment across the lifespan* (pp. 167–196). Washington, DC: American Psychological Association.
- Vandell, P. L., Reisner, E. R., & Pierce, K. M. (2007). *Outcomes linked to high-quality after school programs: Longitudinal findings from the study of promising after school programs*. Philadelphia: Policy Study Associates.

- Venzen, M. A. (2011). *Impact of the Twenty-First Century Afterschool Program on Student Achievement in Mathematics and Language Arts*. 789 East Eisenhower Parkway, PO Box 1346, Ann Arbor, MI 48106: ProQuest LLC.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48.
- Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125.
- Weisman, S. A., Womer, S. C., Kellstrong, M., Bryner, S., Kahler, A., Slocum, L. A., et al. (2003). *Maryland After-School Community Grant Program Part 1: Report on the 2002–2003 school year evaluation of the phase 3 after-school programs*. College Park, MD: University of Maryland.
- Welsh, M. E., Russell, C. A., Williams, I., Reisner, E. R., & White, R. N. (2002). *Promoting learning and school attendance through after school programs: Student level changes in education across TASC's first three years*. Washington, D.C.: Policy Studies Associates Inc.
- What Works Clearinghouse. (2014). *What Works Clearinghouse procedures and standards handbook*. Version, 3, 1–91.
- Wright, R., John, L., Ellenbogen, S., Offord, D. R., Duku, E. K., & Rowe, W. (2006). Effect of a structured arts program on the psychosocial functioning of youth from low-income communities: Findings from a Canadian longitudinal study. *The Journal of Early Adolescence*, 26(2), 186–205.
- Yohalem, N., & Wilson-Ahlstrom, A. (2010). Inside the black box: Assessing and improving quality in youth programs. *American Journal of Community Psychology*, 45, 350–357.
- Zief, S. G., Lauver, S., & Maynard, R. A. (2006). Impacts of afterschool programs on student outcomes: A systematic review. *Campbell Systematic Reviews*, 2, 1–51. <https://doi.org/10.4073/csr.2006.3>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Christensen, K. M., Kremer, K. P., Poon, C. Y. S., & Rhodes, J. E. (2023). A meta-analysis of the effects of after-school programmes among youth with marginalized identities. *Journal of Community & Applied Social Psychology*, 33(4), 882–913. <https://doi.org/10.1002/casp.2681>