A hard outcome for positive youth development programs

Barton J. Hirsch

Northwestern University

For helpful comments on earlier drafts, I am grateful to Nancy Deutsch, David DuBois, David Figlio, Denise Gottfredson, Stephen Hamilton, Reed Larson, Richard Lerner, Milbrey McLaughlin, Karen Pittman, and Patrick Tolan.
A hard outcome for positive youth development programs

There are a multitude of theoretical frameworks and methodological approaches that have come to characterize research on positive youth development (PYD) in after-school programs. Rich analyses of implementation and of how young people navigate social systems, as well as the increasing use of longitudinal research and of randomization procedures, are making marked contributions to an applied developmental science (Hirsch, 2015; Larson, Jensen, Kang, Griffith, & Rompala, 2012; Lerner, Lerner, Bowers, & Geldhof, 2015; McLaughlin, in press). Beyond contributions to science and program design, the findings that result from this research provide evidence that can be used to argue for funding from government and private sector sources.

Yet, not everyone is persuaded by that research. Different types of limitations have been expressed, some theoretical and some methodological. One very strong objection centers around what is referred to as ‘hard outcomes.’ I discuss what is meant by that term extensively later in the paper, but for the moment let us think of hard outcomes as those that are highly meaningful and definitively measured, often by a binary variable, such as whether one has graduated from school, been arrested, or died from coronary heart disease. The argument from those research critiques, which can easily work their way into disagreements over funding and policy priorities, is that PYD has not demonstrated impact on hard outcomes and, indeed, that the field cannot even identify a relevant hard outcome.
Most of us do not believe that the merit of the field is reducible to whether we have demonstrated an impact on a hard outcome. Many believe that the processes and outcomes we have reported provide the core of an alternative framework that is being developed that can speak well to merit. There are desirable outcomes beyond those that can be conclusively measured at present. Yet we fail to engage the hard outcome argument at our peril. It has the allegiance of important figures and, in many fields, when the stakes get high and the argument heated, hard outcomes can come into play. In this paper, I take on the hard outcome objection and, based on research conducted by my research group, argue that PYD has at least one hard outcome.

The first part of the paper considers the concept of a hard outcome. In order to develop some shared understandings, I begin by considering outcomes from after-school, or extended education, programs focused on academics and then consider PYD outcomes. I mine these examples to develop broader principles about what makes an outcome a hard outcome. This approach draws on what philosophers do in analyzing a concept, and has similarities as well to how qualitative researchers discuss the broader implications of their data.

The second part of the paper considers how these principles apply to PYD programs. Limitations of some of our variables and measures are noted. I then turn to some of my own studies and argue that we can usefully locate a hard outcome with respect to job readiness, an important developmental objective that is crucial to making
a successful transition to adulthood. I focus particularly on studies of minority youth, who encounter considerable difficulty in obtaining employment. A labor market outcome is likely to be of broad appeal, assessment advances have produced a well-defined measurement, and the kind of soft skills that can lead to positive hiring decisions are frequently the target of PYD programs. I conclude by suggesting what we need to do in the future to buttress our position on hard outcomes.

**The rationale for academic programs**

Outcomes for academically oriented after-school programs seem straightforward when the program has a clear focus on improving student abilities to succeed in school. The typical program involves tutoring around class concepts and homework assignments. Outcomes might include better attendance, school grades, scores on standardized tests and, taking the long view, fewer drop-outs and higher rates of graduation from high school (secondary school). There is a clear consensus that all of these are good outcomes. For students at academic risk, one might question whether the first set of outcomes are intermediate outcomes and whether the latter outcomes are what really matter. An argument could be made that if a program impacted the longer-term outcomes, then the results would be more impressive than demonstrating shorter-term impacts. Presumably that would be because the latter outcomes are ultimately of greater importance. Methodologically, most of these outcomes are easily measured: there are shared criteria for whether a student is either absent or not on a
particular day and a high school diploma is either received or not. Grades awarded by a teacher have a more subjective element, which is part of what drives the increased emphasis on standardized, and presumably more objective, tests.

The situation is less clear-cut when we consider programs that are designed to provide academic enrichment beyond that provided by the school curriculum. Let us consider project-based learning programs in science, which aim to foster a deeper learning than is available in traditional science curricula. There are a number of such offerings in after-school or extended education settings. In these programs, students gain knowledge and skills by assuming the role of scientists investigating real-world problems. Students learn by doing, rather than by memorization. They grapple with understanding the underlying scientific principles through conducting experiments, articulating and defending their views, applying insights to different situations, and revising their ideas in the light of new evidence. Project-based learning activities take place in some school science classes; the advantages of such programs in after-school settings are that they permit considerably more time to be spent on a project than is available during a typical school class, enabling a more in-depth immersion in the topic.

Despite its popularity in university circles, such programs have not been greeted with unfailing enthusiasm among legislators or the wider public. Indeed, in the USA, there has been a political and popular backlash against the Common Core, a major reform initiative which seeks such deeper understandings across a range of subjects.
Some question the value of taking time away from traditional curricula and wonder whether immersion in new approaches to learning might be a less effective route to improving test scores, getting accepted into competitive colleges, or obtaining better grades in traditional college science courses. Disagreements about project-based learning programs may reflect differences in values and goals for science learning, as well as uncertain knowledge about the relation between program participation and outcomes (Duschl, Schweingruber, & Shouse, 2007).

These objections apply especially to science classes during the regular school day, when project-based learning modules take away time that might otherwise be devoted to the traditional curriculum. But the tensions they highlight bear on after-school programs as well. Let me provide an example from my own experience. I served as a consultant to an interesting and engaging project-based science program in a comprehensive after-school center. The students were engaged and learned a lot. But it was the only academic-oriented program that was offered. The young people, and their parents, could not choose to take another program. Given that many of the youth did not get good grades in school, there was a question of whether some of the parents might have preferred a tutoring program that focused on more traditional academic content and that had more obvious ties to school classes and grades.

There are also measurement concerns with respect to a project-based approach. Test developers are still at an early stage of developing instruments to assess deeper
learning. Furthermore, one might legitimately question whether paper-and-pencil tests are the best way to assess more complex understandings of scientific phenomena and the ability to engage more meaningfully in the practices of science. Thus, there might be both disagreement about values and goals, as well as uncertainty about valid measurement.

**The rationale for PYD programs**

PYD programs are oriented to youth strengths and seeing young people as resources to be developed (Benson, Mannes, Pittman, & Ferber, 2004: Lerner et al., 2015). A shorthand way often used to capture the qualities that such programs aim to develop is to refer to the 5 Cs: competence, confidence, character, connection, and caring. Program activities should be challenging, framed in terms of strengths rather than deficits, and ideally involve the assumption of leadership or adult-like roles. For example, Boys & Girls Clubs of America, one of the largest programs in the USA, offers in-club stores where youth learn to sell merchandise (refreshments, sweatshirts) and do cash transactions; arts oriented activities such as dance teams; sports instruction, teams, and leagues; and social skill and resiliency building programs, such as Smart Girls (Hirsch, 2005; Hirsch, Deutsch, & DuBois, 2011).

All of these appear to be worthwhile goals. They tend to resemble academic enrichment activities more so than academic tutoring activities; indeed, PYD programs are sometimes labelled as enrichment. They do not have an obvious, external outcome,
like grades or graduation. At times PYD programs try to trade on a presumed relation to those academic outcomes, but that has been termed a “big lie” (Halpern, 2006), as it may be difficult to achieve specific academic outcomes unless they are explicitly targeted (Durlak, Weissberg, & Pachan, 2012) and they are often not targeted in PYD programs. There is not the clear political and popular consensus for funding PYD programs as opposed to those that target better defined, more understandable and instrumental goals.

There are also measurement problems in assessing PYD outcomes. The outcomes are not specified in a tangible way like graduation. They are conceptually harder to pin down in terms of specific behavioral manifestations. Assessment is frequently via self-report, which is not considered objective. Moreover, a continuous scale rather than binary score is typically utilized. But what is the real difference, say, between a score of 2 and 3 on a questionnaire? Assessment of PYD concepts, thus, does not currently provide the basis for a broad consensus in support of a particular measurement approach.

**What is a hard outcome?**

The term ‘hard outcome’ is frequently used in USA policy discussions among not only professionals, but also the broader public through newspaper and magazine articles. Nonetheless, you can search dictionaries or Google and not find an analysis of what is meant by a hard outcome. Or at least that was the case with me. One policy
researcher I spoke with compared it to the famous expression that United States Supreme Court Justice Potter Stewart used in 1964 to describe his threshold test for obscenity: “I know it when I see it.” Nevertheless, we shall venture into the lion’s den and seek to unpack the notion of a hard outcome. Let us begin by using the examples we have discussed regarding academic and PYD after-school programs.

In my view, the strongest outcomes discussed earlier were both meaningful and measurable and I propose that we use those criteria to define what constitutes a hard outcome. Each of these terms needs its own elaboration.

In the policy world, decision makers prefer to fund programs that address goals that they consider meaningful. They are likely to want there to be a high level of consensus about the goal’s meaningfulness among a relevant interest group or the population at large. In the absence of such consensus, the policy maker will want to have considerable personal confidence in its meaningfulness and perhaps some sense that, if properly communicated, a consensus would likely emerge. Meaningfulness also includes the sense that it is worth spending public funds to achieve this goal.

In our examples, graduating from high school or college would be considered meaningful goals. There is not presently a consensus that acquiring “deep” understanding of scientific phenomenon is preferable to acquiring the kinds of understanding that occur in typical science classrooms. In medicine, to consider another field that considers hard outcomes, decreasing deaths due to heart attacks or stroke is
considered meaningful, whereas decreasing ordinary cuts and bruises would not be considered meaningful.

For PYD, it is not clear that any of the outcomes identified earlier have this consensus level of meaningfulness. For example, during a conference on poverty, a legislator from one of the USA’s most prominent liberal families told me that he would not vote to fund PYD outcomes when such funds as were available needed to be spent on “actual” community problems, such as drug abuse or violence. Reducing rates of the latter two were meaningful outcomes that enjoy a strong political and public consensus. There is not the same level of interest around achieving PYD outcomes (cf. Benson, 2006).

Besides meaningfulness, the goals need to be measurable. Measurable goals, in relation to hard outcomes, typically are thought to be objective. They also tend to be readily understandable. Some goals of this sort can be measured on a continuous scale, such as lifetime income, but that is a ratio scale of a concrete, tangible unit, which helps to make it understandable. Such scales are relatively rare in social science. More commonly, measures of the most understandable goals rely on a discrete classification system (sometimes continuous scores are converted into a classification system). Often the classification is binary, which is the most basic classification available, such that an individual either has it or they do not have it. For example, an academic course has been passed or not. High school graduation has been achieved or not. A person has been
diagnosed as having had a heart attack or does not receive that diagnosis. A person is either dead or alive. This simple, clear-cut distinction facilitates understanding and provides the best foundation for a consensus that the measure is or is not appropriate.

Applying this binary classification to common PYD concepts—is a person morally praiseworthy or not, is the person resilient or not, is he or she civically engaged or not—may work for individuals who are at one extreme or the other, but seems less certain for all of those who are not at the extreme. Moreover, the concepts may not be stable across domains; we have seen well-known religious or political figures who have slipped badly in their interpersonal relationships.

PYD concepts could be linked to non-PYD, binary criteria, but one needs to be careful about this. For example, one might theorize that increasing PYD leads to lower crime rates or teen pregnancy. Each of the latter can be measured using binary variables. But a key tenet of PYD is that it is about how healthy and effective development is achieved and not just how problems are reduced. Indeed, for a movement that emphasizes the value of the positive, it would be grossly inconsistent to point to lower rates of problem behavior as the ultimate justification. It is difficult to imagine that PYD advocates would want to assume this position. An external outcome may be highly useful, but it would need to be framed in ways consistent with a PYD philosophy.
Ordinary language philosophy (Austin, 1962; Searle, 1969) alerts us to other factors to consider when claims regarding hard outcomes are discussed. This perspective on language draws our attention to what one may be trying to accomplish when making claims about hard outcomes. The first thing that might be noticed about hard outcomes is that ‘hard’ is often used to refer to or connote qualities of toughness and aggressiveness. Indeed, this is often the case when one claims to have a hard outcome or when one challenges another to demonstrate a hard outcome. The discussion becomes serious, potentially verbally combative, and aimed at reaching a clear resolution of the issue. If I claim that my program has demonstrated that it had a hard outcome, I am making a bold assertion and must be willing to back it up with clear evidence. Similarly, if I assert that a program does not have a hard outcome, I am challenging the meaningfulness or measurement of its goals. Consideration of whether a program has a hard outcome means subjecting it to a very strong test of its value. For PYD, a field that has ambitious aspirations, that considers itself a major institution of child and youth development (Halpern, 2006), a hard outcome is extremely desirable, if not ultimately necessary. It confers a legitimacy that is difficult to dispute.

Yet, according to this analysis of the concept of hard outcome, there does not presently exist a hard outcome for PYD programs. There is no clear consensus as to the meaningfulness and long-term manifestations of many PYD terms. PYD concepts can have a certain sponginess to them, so that they can be difficult to pin down, specify, and
measure. It is awkward, if not impossible, to reduce many PYD variables to binary PYD outcomes, which problematizes the ability to develop consensual, easily understood measures.

I believe, as do many others, that PYD outcomes are important. We certainly want a society in which moral, artistic, and civic endeavors, for instance, are valued, even if they may not have obvious hard outcome indicators. But I also accept that PYD programs would have a broader and stronger foundation, and could repel attacks against it more easily, if they could point to a hard outcome. Given the obstacles to developing such an outcome using current core PYD terms, I propose that we look at an alternative outcome—consistent with a PYD approach—that can be achieved using PYD programs. I consider one such possibility in the next section and others in the conclusion.

A proposed hard outcome for PYD programs

My proposal for looking to the labor market for a hard outcome grows out of evaluation research that my group conducted on a major after-school program. After School Matters, located in Chicago, is widely considered the flagship program in the USA for after-school PYD programs that focus on youth in high school (grades 9-12, typically ages 14-18). It provides apprenticeship-like training for 180 hours each academic year (9 hours/week X 20 weeks) in a wide variety of content areas, including the arts (e.g., dance, song writing), technology (e.g, web design), and sports (e.g.,
teaching high schoolers to be counselors to young children in summer camps). At the
time of the research, the organization’s primary outcomes were PYD and marketable job
skills. Almost all of the youth in the evaluation were minorities from low-income families
(African American = 77%, Hispanic = 23%, with 92% in low-income families).

In stepping back and considering broader implications of this research in a recent
book (Hirsch, 2015), I framed the findings regarding marketable job skills in terms of
minority youth employment. Youth workforce development is an important objective for
every society. Young people are at a disadvantage in securing employment and, in the
USA, the unemployment rate for youth is often approximately twice the overall
unemployment rate. Among youth, those from a minority background have the highest
rate of unemployment, typically twice the overall youth unemployment rate. In the USA,
this would primarily be African Americans (blacks) and Hispanics. In Chicago, for
example, minority youth unemployment is a major problem, with 41% of black youth
out of work and out of school (Cordova & Wilson, 2016). The situation is not so
dissimilar in Europe. In a cross-national meta-analysis of field experiments, France,
Sweden, Great Britain, and Belgium all had higher rates of hiring discrimination against
racial or ethnic minority groups than did the USA (Quillan et al., 2016). Over the long
term, failure to incorporate minority youth into the workforce can tear at the social
fabric of society. In the USA, increasing the job readiness of minority youth is a social
and economic outcome that would find support across the political spectrum.
One of our research tasks in the After School Matters evaluation was to develop a measure of marketable job skills or job readiness. To address this need, we worked with senior human resource (HR) professionals to develop a mock job interview (the Northwestern Mock Job Interview is reprinted in Hirsch, 2015). In the business world, HR workers help to develop hiring criteria and are often actively involved in the job interview process. By collaborating with HR professionals, we made sure that the mock job interview made business sense, as job interviews are a crucial part of the hiring process in both the USA and Europe, and probably in many other countries as well (Dipboye, Macan, & Shahani-Denning, 2012; Macan, 2009; Rosenbaum, 2001). Interviews are used to make judgements as to job skills and fit with the culture of the organization. As I will discuss shortly, these judgements often reference personal and social skills, thus providing a link to PYD programs. We had HR professionals conduct the mock job interviews so that the findings would have credibility in the business and policy worlds. The hiring decision was measured on a 5-point scale in which each level of the scale was anchored by a description that corresponded to how HR professionals rate applicants. Ratings of 4 (would hire the applicant) and 5 (would hire with enthusiasm) were deemed “hired” and ratings of 1, 2, and 3 were deemed “not hired.” Thus, we had the ability to convert this assessment into a binary variable. Moreover, being hired or not hired is a classification that is easily understood by all adults.
Would a rating of being hired in a mock job interview constitute a hard outcome? An increase in the level of job skills among minority youth, their job readiness, is an important, consensual goal. On the measurement side, there would be a consensus that the outcome was understandable and that a job interview was the appropriate measurement procedure. There would be a consensus that HR professionals were suited to make such an assessment. Some judgment factors into the hiring decision, so that it is not completely objective (we used semi-structured interviews, with scores of 1, 3, and 5 for each job skill item anchored by descriptions, to enhance inter-rater reliability). However, judgement is an important element in how hires are made in real life and employers would not want it to be otherwise.

The major stumbling block is whether ratings of hirability in a mock job interview would be the hard outcome, or whether such a status should be reserved for hiring rates for real jobs. In terms of longer-term or ultimate outcomes, getting an actual job would be the clearer and most convincing result. However, many young people will not get actual jobs because there are not enough jobs available or because those who are older and have more job experience will be hired instead of them. Moreover, a meta-analysis of field experiments point toward the persistence of job discrimination against adults who are members of ethnic and minority groups (Quillan et al., in press) and it is reasonable to consider that such discrimination might be operative with respect to young people as well. Thus, failure to obtain a job need not reflect that a young person
is not qualified for work. We faced this issue in training the HR professionals in how to use the mock job interview. We instructed them to rate whether the young person was hirable on the assumption that a job would be available for all applicants rated as hired.

A hiring rating on a mock job interview should be seen as a type of certification. The award of a high school diploma certifies that the student has passed various courses and examinations, administered by the appropriate professionals in that field. Similarly, a positive hiring decision by a HR professional attests that the young person has acquired sufficient job skills to be ready for employment, as determined by HR professionals, who are gate-keepers regarding employment, the ones who can determine whether a young person enters the workforce. It is not realistic, and therefore not appropriate, to have actual employment rates be the measure of job readiness for youth given the reasons specified earlier. Thus, a positive hiring decision in a mock job interview, administered by an HR professional or other suitable person, should be considered a hard outcome of job readiness for young people.

**Linking Mock Hiring Rates to PYD Programs**

The After School Matters research yielded a number of findings that tapped the importance of soft skills to interviewer hiring ratings. Soft skills are sometimes referred to as 21st century skills or noncognitive skills (Murnane & Levy, 1996; Pellegrino & Hilton, 2012; Schanzenbach, Nunn, Bauer, Mumford, & Breitwieser, 2016). These generic skills—such as teamwork, communication, leadership, problem solving, initiative,
and self-regulation—are useful in a wide variety of jobs. They are also skills that are acquired in many good PYD and extended education programs.

The soft skill findings emerged from both quantitative and qualitative analyses that were conducted as part of our randomized controlled trial (Hirsch, 2015). The two mock job items with the strongest correlations with hiring, across both the intervention and control groups, were ratings of the young person’s maturity and communication skills. A good number of other soft skills significantly correlated with mock hiring. In terms of qualitative data, debriefings of HR professionals after each day’s interviews indicated that they were most concerned with youth soft skills. They asserted that businesses preferred to teach hard skills once on the job.

We also conducted an extreme-groups analysis in which we contrasted observations of the two After School Matters apprenticeships with the best hiring rates (vs. their control groups) with those of the two After School Matters apprenticeships with the worst hiring rates (vs. their control groups). In these qualitative analyses, the quality of teamwork and communication stood out as the two most important differences between the best- and worse-hire apprenticeships (Hirsch, 2015, ch. 3).

There are quite a few other examples of soft skills reported in intensive, qualitative research on PYD after-school programs. These range from the groundbreaking volume by Milbrey McLaughlin and colleagues (1996) on the development of a culture of hard work, responsibility, loyalty, and participatory
decision-making; to Reed Larson’s work on teamwork, initiative, and strategic thinking (Larson, 2000, 2007; Larson & Angus. 2011); to our own group’s research (presented earlier in the chapter) on personal and social skills developed in Boys & Girls Clubs of America (Hirsch et al., 2011). It should be noted that these publications do not explicitly refer to soft skills, but the skills they describe correspond well to those that are labelled as soft skills in the workplace. Of course, not all PYD programs are implemented well and weak programs may well not help youth develop these skills.

Would learning soft skills by themselves lead to better youth hiring rates? It became clear in our evaluation research that although the HR interviewers believed that many of the young people had skills and experiences that employers value, too many youth did not know that they had those qualifications or did not communicate them successfully in the interview. At the request of high-level administrators in the public school system in Chicago, we developed a structured, 5- or 6-session program (45 minutes/session) to teach job interview skills to 11th grade youth in vocational classes. The curriculum focused on teaching the students how to communicate their soft skills effectively to HR professionals. The students’ regular vocational teacher received professional development sessions from us and then implemented the curriculum. Evaluation of the program, using pre-post mock job interviews (n = 75 students across 5 classes in 4 schools) administered by HR professionals, but no control group, indicated a doubling of the mock hiring rate (Hirsch, 2015). These findings suggest that training in
communication skills specific to job interviews should be an important part of programs that seek to impact job readiness.

Based on its initial implementation, we knew that the interview training could be integrated into vocational classes in school (for a discussion of how job interview training might lead to a reimagining of how high schools can educate students for both academics and careers, see Hirsch, 2017). The next step was to explore how well it might be implemented as part of an existing after-school or extended education program. Integrating it into such programs could potentially enhance the reach and effectiveness of each program. It could do so, in part, by teaching young people to communicate effectively how the skills they have learned in an after-school program could enable them to be successful in a particular job. For example, youth in a project-based science program will typically work together as a team and teamwork skills are highly valued in today’s workforce.

Accordingly, a member of our staff taught the program in a dance apprenticeship program run by After School Matters and it fit quite well into the curriculum. We relished the opportunity to incorporate the interviewing program into a performance arts program, as dance involves a high level of teamwork, which is a core job skill. The young people learned how to talk about their dance experiences in ways that would make sense to job interviewers. Moreover, observing our program, the two dance instructors shared information with the students about nonperformance jobs in arts
organizations that they had never mentioned before, sparking interest among students who would not meet the criteria for being hired into performance jobs.

We believe that many PYD staff would be able to implement the program successfully, but not all. In particular, some arts instructors might be philosophically opposed to introducing the prospect of extrinsic rewards into their program. This could be a problem in a single-focus, PYD program with a very small staff. On the other hand, in organizations with multiple staff, it is likely that one or more of them would have the interest and personal skill prerequisites to be able to learn the program and implement it well (training issues are discussed in Hirsch, 2015). As many PYD staff do not have experience with structured programs, it might be necessary to modify the program to make it more consonant with their backgrounds (cf. Hirsch, 2005, Ch. 7), but this remains to be determined. It is likelier that staff in academically-oriented programs would have had prior exposure to structured curricula, given their greater orientation to academic content (in which structured curricular are not uncommon), and this should ease their ability to implement the program successfully.

**Conclusions and Future Directions**

Our ability to successfully communicate the strengths of PYD programs to policy makers and funders has been constrained by the lack of a hard outcome. Potential hard outcomes that are intrinsic to our conceptualization of PYD are not readily definable or measurable in ways that might elicit consensus agreement. Using an extrinsic, negative
outcome, such as a decline in problem behaviors, although useful, seems inconsistent as an ultimate justification with the PYD ethos. Instead, I have argued that locating a hard outcome with reference to the labor market makes practical and conceptual sense. In particular, improving the job readiness of minority (and low-income) youth is an aspect of PYD that can make sense in the policy world. In the USA, at least, it would not be possible in a policy framework to guarantee equity in outcomes (e.g., a job guarantee), but there is support for achieving equity in opportunity, for providing a sound educational foundation from which to seek and gain employment. In an age of fiscal austerity, the job readiness argument may carry some weight in funding deliberations.

Job readiness is an important hard outcome and developmental marker. It certifies that a young person has the skills and experiences that qualify them for entry into the work world. More research, though, is clearly needed. Studies that demonstrate a linkage of mock interview hiring with actual hiring rates would be valuable, though the strength of the association would be constrained by previously identified factors. One could also examine the linkage of performance on the mock job interview to ratings by subsequent supervisors, perhaps in summer internships or jobs. To extend the utilization of mock interviews, research is needed on whether non-HR interviewers could do an adequate job in making job skill assessments and hiring ratings. This research is important as many PYD programs will not be able to pay HR interviewers for these interviews and there may not be sufficient interest among HR
professionals to conduct a very large number of interviews. Training programs for non-HR interviewers need to be developed and a satisfactory level of inter-rater reliability would need to be obtained between HR and non-HR raters.

The field of PYD should not rely exclusively on mock job hiring for a hard outcome. The more hard outcomes that can be articulated, the stronger the potential of the field. For example, a meta-analysis of after-school programs designed to impact physical fitness (Beets, Beighle, Erwin, & Huberty, 2009) referenced use of a number of biomarkers as outcomes, including body composition (e.g., percentage body fat), physical fitness (e.g., cardiovascular fitness), skeletal health (e.g., bone mineral density), and blood lipids (e.g., cholesterol). There are a number of popular PYD programs—sports activities, dance—that could be expected to impact those measures. Cortisol is another biomarker. Increasingly used to assess stress level, this could be a relevant outcome for resiliency-oriented programs. There is a medical literature on conceptualizing surrogate endpoints (intermediate outcomes, such as cholesterol levels) and clinical endpoints (final outcomes, such as heart attack) which can help PYD researchers think through some of the theoretical and measurement issues (Biomarkers Definitions Working Group, 2001; Strimbu & Tavel, 2010).

Megan Mekinda and I (Hirsch, 2015, ch. 4) have discussed how assessments can be made of the skillfulness of group activities using criterion-referenced outcomes. One of the best After School Matters apprenticeships used expert judges to make
multidimensional ratings of the skillfulness of food dishes created by competing culinary teams. We demonstrated how the type of assessment utilized in the culinary assessment, made by both professionals in the field and informed consumers, could be adapted to artistic and technology oriented PYD programs. It may also be possible to extend this assessment to individuals. This has some similarity to expert judgements made in science fairs. The higher the credibility of the expert judges, the greater the confidence there will be in this type of assessment.

A rather different approach might seek to identify hard outcomes from international conventions that have some level of legal authority. I am thinking here of the UN Convention on the Rights of the Child (1989). The USA is not a signatory to the convention; that convention has no force in the USA and, indeed, is little known even among developmental scientists. However, almost all other countries are signatories to it. Articles 3, 4, 12, 13, 17, 23, 24, 27, 28, 29, and 31 of the convention provide support for PYD programs, although being a signatory to the convention has not prevented a marked decline in youth work in some countries, such as the UK (Jeffs, 2015).

If PYD researchers conduct long-term follow-ups, data in relation to traditional hard outcome measures may become available as well. McLaughlin (in press), for instance, studied long-term adult outcomes of those who participated in a comprehensive PYD program when young. The program was located in what was habitually described as one of the most “notorious” public housing projects in the USA.
Despite the adversity and disadvantage the young people faced, an unusually large number of them went on to obtain graduate degrees (including doctorates). Although there was no control group, the outcomes are well beyond what might be expected.

PYD has been carving a rich place for itself in the world of programs, as well as theory and research. I have argued that we should broaden our efforts to consider potentially relevant hard outcomes. We should not become obsessed with hard outcomes, but neither should we ignore them. Our field will mature if it debates these issues and develops new understandings from the challenging work that this effort will entail.
References


Larson, R., & Angus, R. (2011). Adolescents’ development of skills for agency in youth


