

Implementation Research in Impact Studies of Interventions

CRT Summer Institute

Northwestern University, July 14, 2025

Carolyn J. Hill (MDRC)

This session

- What is implementation research? Why should you care about it?
- Introduce resource
[*Conducting Implementation Research in Impact Studies of Education Interventions: A Guide for Researchers*](#)
- Overview of key ideas and questions about
 - Intervention components
 - Intervention contrast
 - Context features

What is implementation research?

- Analyzes whether and how policies, practices, procedures, and interventions are put into use by frontline staff, managers, policymakers, and other actors
- Analyzes strategies that actors use, contexts in which implementation occurs, and facilitators and barriers to implementation
- *When conducted as part of an impact study, involves*
 - Detailed understanding of the intervention that is planned *and* the intervention that is actually implemented
 - Analytic thinking and use of frameworks and theories
 - Advance planning and anticipation of different scenarios
 - Careful alignment between key constructs and measures
 - Both quantitative and qualitative methods

Why should you care about implementation research in an impact study?

“It is the plunge into the cold bath of implementation that has given evaluation the chills.”

Angela Browne and Aaron Wildavsky, 1983, “What Should Evaluation Mean to Implementation,” Chapter 9 in Pressman & Wildavsky, *Implementation, 3rd edition* (University of California Press, 1984).

Why should you care about implementation research in an impact study?

Well-planned and executed IR can

- Provide information for interpreting the study's impact findings beyond an up-or-down conclusion about whether the intervention “worked” or “didn't work”
- Inform future research and practice, regardless of whether estimated impacts are positive, negative, or null
- Inform future implementers about whether they might want to adopt or adapt the intervention tested in the study

Standards for Excellence in Education Research (SEER)

<https://ies.ed.gov/use-work/standards-excellence-education-research-seer>

Text from July 2024:

STANDARDS FOR EXCELLENCE IN EDUCATION RESEARCH

- About
- Pre-registration
- Open data
- Equity
- Components
- Implementation
- Cost analysis
- Outcomes
- Generalization
- Scaling

About

Rigorous education research that is transparent, actionable, and focused on consequential outcomes has the potential to dramatically improve student achievement. Since 2002, the Institute of Education Sciences has supported rigorous evidence-building about education policy, programs, and practices. SEER complements the WWC's focus on internal validity by supporting a distinct, IES-wide effort that emphasizes additional factors that can make research transformational.

IES Director Mark Schneider introduced SEER in two blogs. You can read those blogs [here](#) and [here](#). IES has iterated on SEER frequently and expects to continue to do so as we receive feedback from researchers, practitioners, and policymakers. IES' goal is to produce concrete standards against which research can be judged to award different levels of certification. Following the well-known [LEED](#) levels of certification these include: Certified, Silver, Gold, or Platinum.

To provide feedback, please contact NCEE.Feedback@ed.gov.

SEER
implementation
standard and
resources

<https://ies.ed.gov/use-work/standards-excellence-education-research-seer>

Pre-register studies

Make findings, methods, and data open

Equity

Identify interventions' components

Document treatment implementation and contrast

Analyze interventions' costs

High-quality outcome measures

Facilitate generalization of study findings

Support scaling of promising interventions

SEER Standards

- Researchers must document how, and the context within which, the treatment was implemented.
- Researchers must document the counterfactual condition(s), including its context.
- Researchers must measure the essential elements of the treatment contrast between the treatment and control conditions.
- Researchers must measure the fidelity of an intervention's implementation.

Recommendation

- Researchers should document, and identify opportunities to learn from, adaptations of the intervention that were observed during implementation

Resources

- Read the NCEE Publication *Conducting Implementation Research in Impact Studies of Education Interventions: A Guide for Researchers*, by Hill, Scher, Haimson, and Granito (2023).
 - [Conducting implementation research in impact studies of education interventions: A guide for researchers](#) (1.93 MB)
- [View the archived webinar](#) on *Conducting Implementation Research in Impact Studies of Education Interventions: A Guide for Researchers*
- [View the archived webinar](#) on *Prioritizing and Selecting Context Features in Education Impact Studies: A 4R Lenses Approach*, and the webinar slides.
- [Prioritizing and Selecting Context Features in Education Impact Studies: A 4R Lenses Approach](#) (5.71 MB)

What We're Reading

Hamilton, G. & Scrivener, S. (2019). *Measuring treatment contrast in randomized controlled trials*. MDRC Working Paper. Retrieved from:

[Measuring Treatment Contrast in Randomized Controlled Trials](#) (511.19 KB)

Lemons, C. J., Fuchs, D., Gilbert, J. K., & Fuchs, L. S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher*, 43(5), 242–252. <https://doi.org/10.3102%2F0013189X14539189>

Weiss, M., Bloom H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management*, 33(3), 778–808. <https://doi.org/10.1002/pam.21760>

Sections of the implementation guide

Introduction

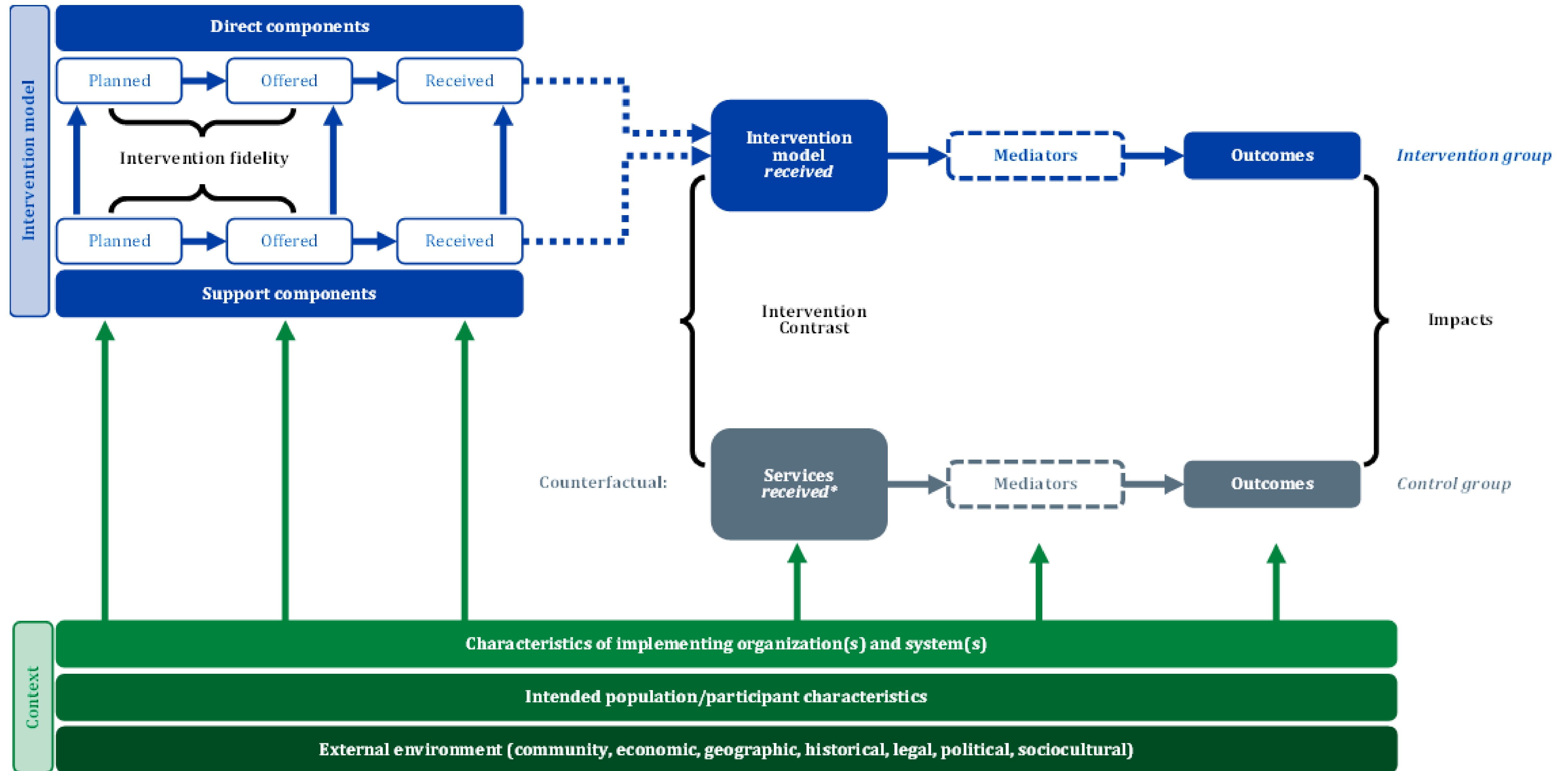
- Section I.** Establish research questions and begin developing a plan for implementation data collection, measurement, and analysis
- Section II.** Specify details about the intervention and its implementation
- Section III.** Specify details about context
- Section IV.** Specify details intervention contrast
- Section V.** Analyze and report details about the intervention and its implementation as part of the impact study

Appendices

Broad IR research questions in an intervention study

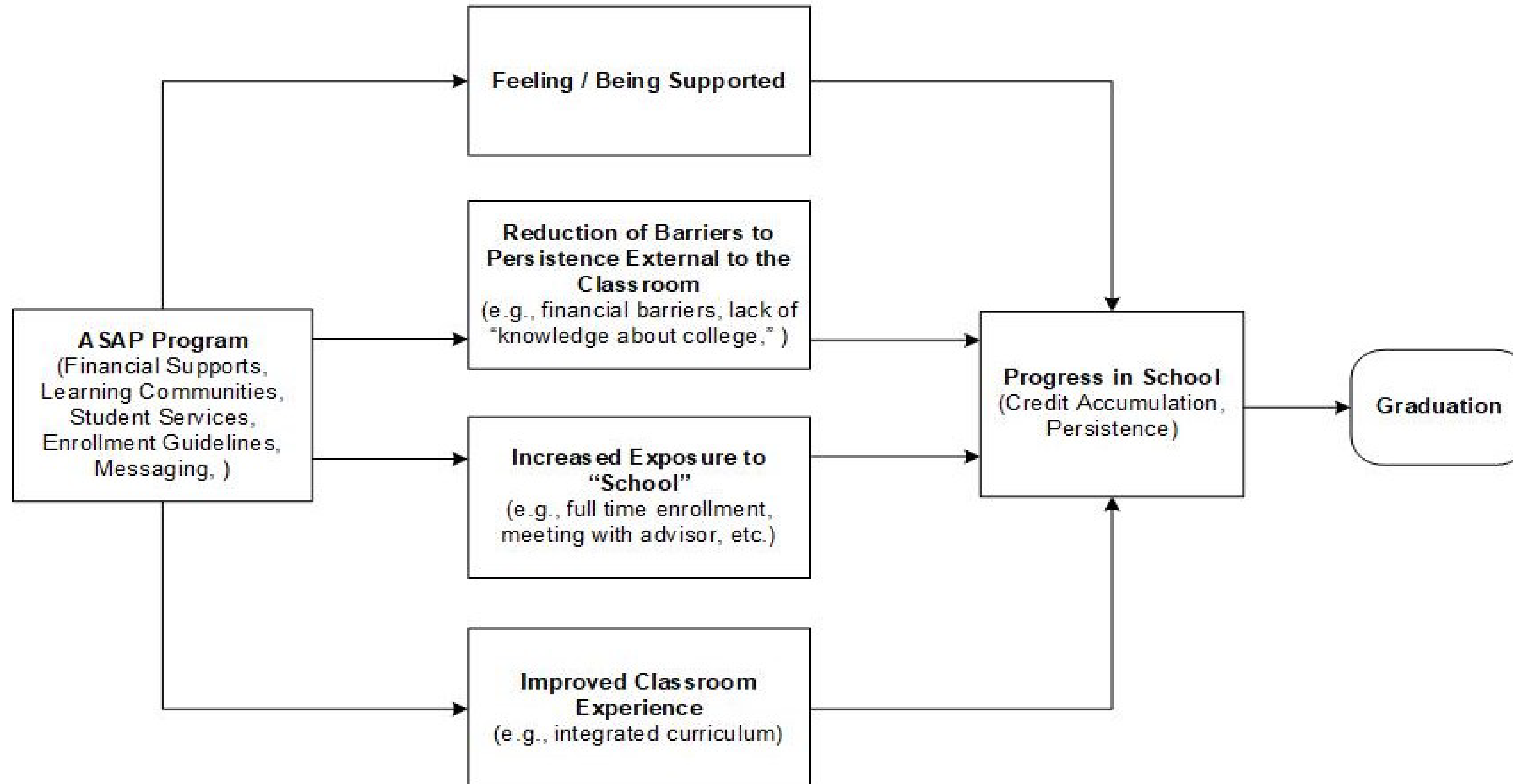
1. What components of the intervention were implemented?
How did implemented components vary?
2. In what contexts was the intervention implemented?
How did contexts vary?
3. What was the intervention contrast?
How did contrast vary?
4. What aspects of implemented intervention components, context, and contrast are associated with (moderate or mediate) impacts?

Exhibit 1. A conceptual framework for linking the intervention model, contrast, and impacts



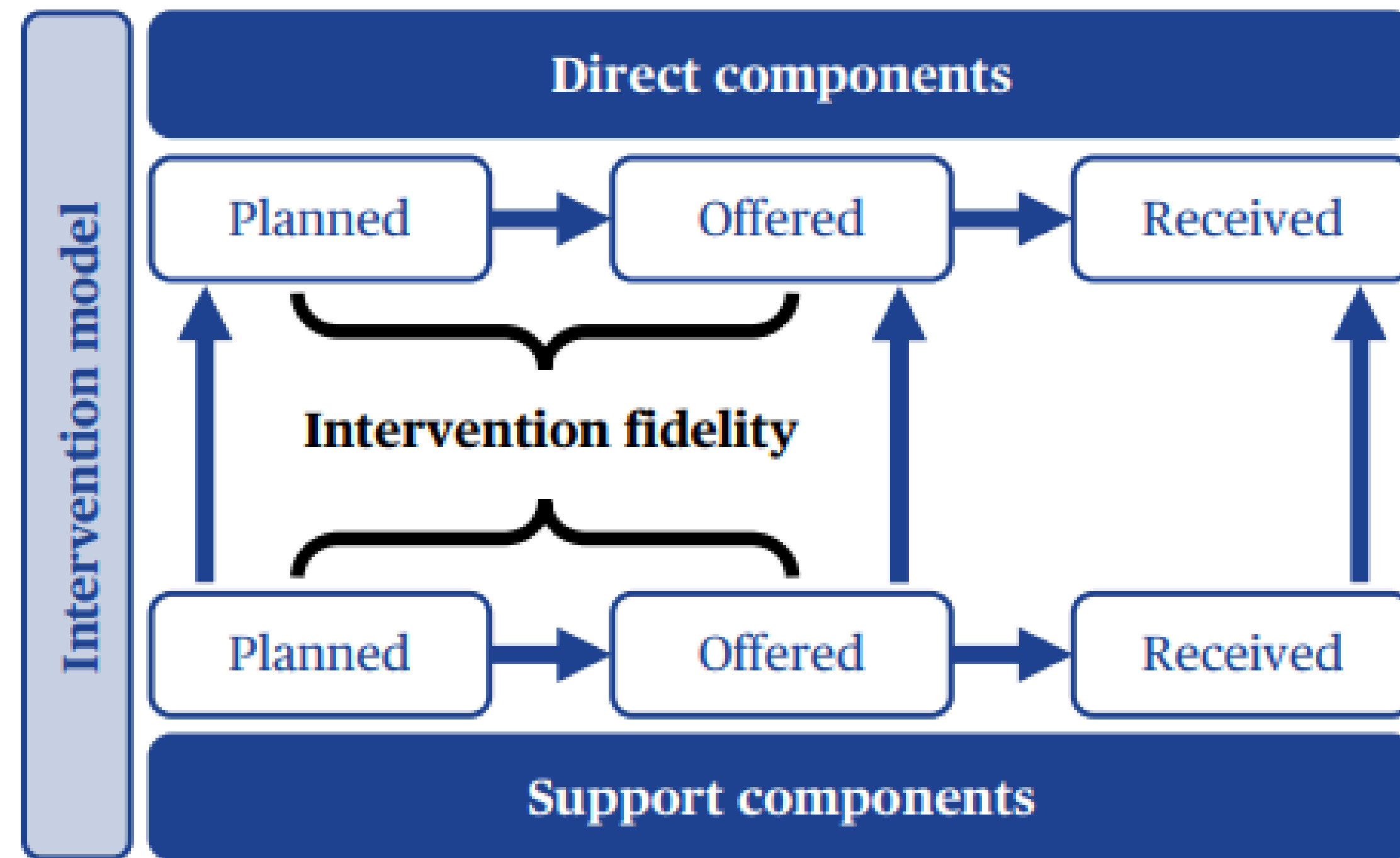
Broad Logic Model for CUNY ASAP

City University of New York, Accelerated Study in Associate Programs



Source: Scrivener et al, MDRC

Intervention Planned, Offered, Received



Four dimensions of *each* planned intervention component

Dimension	Question of interest	Examples
Content	What is the component or service?	Course topics, messaging, \$, advising
Quantity	How much of the component is provided? <i>Frequency:</i> How often? <i>Time:</i> How long is each encounter? <i>Duration:</i> Over what time period? <i>Reach:</i> What % target pop reached?	•1x/week •30 minutes •2 semesters •80%
Mode	How, when, and by whom is the service/intervention provided?	In person, Tuesday afternoons, trained paid adviser
Quality	How well is the component delivered? (What would constitute “high quality” for this component?)	Tailoring to student needs, engagement, ?

Specifying the *planned* intervention components

Direct components

involve direct interaction with the intended population whose impacts are hypothesized to change as a result of the intervention and whose impacts you will measure

Support components

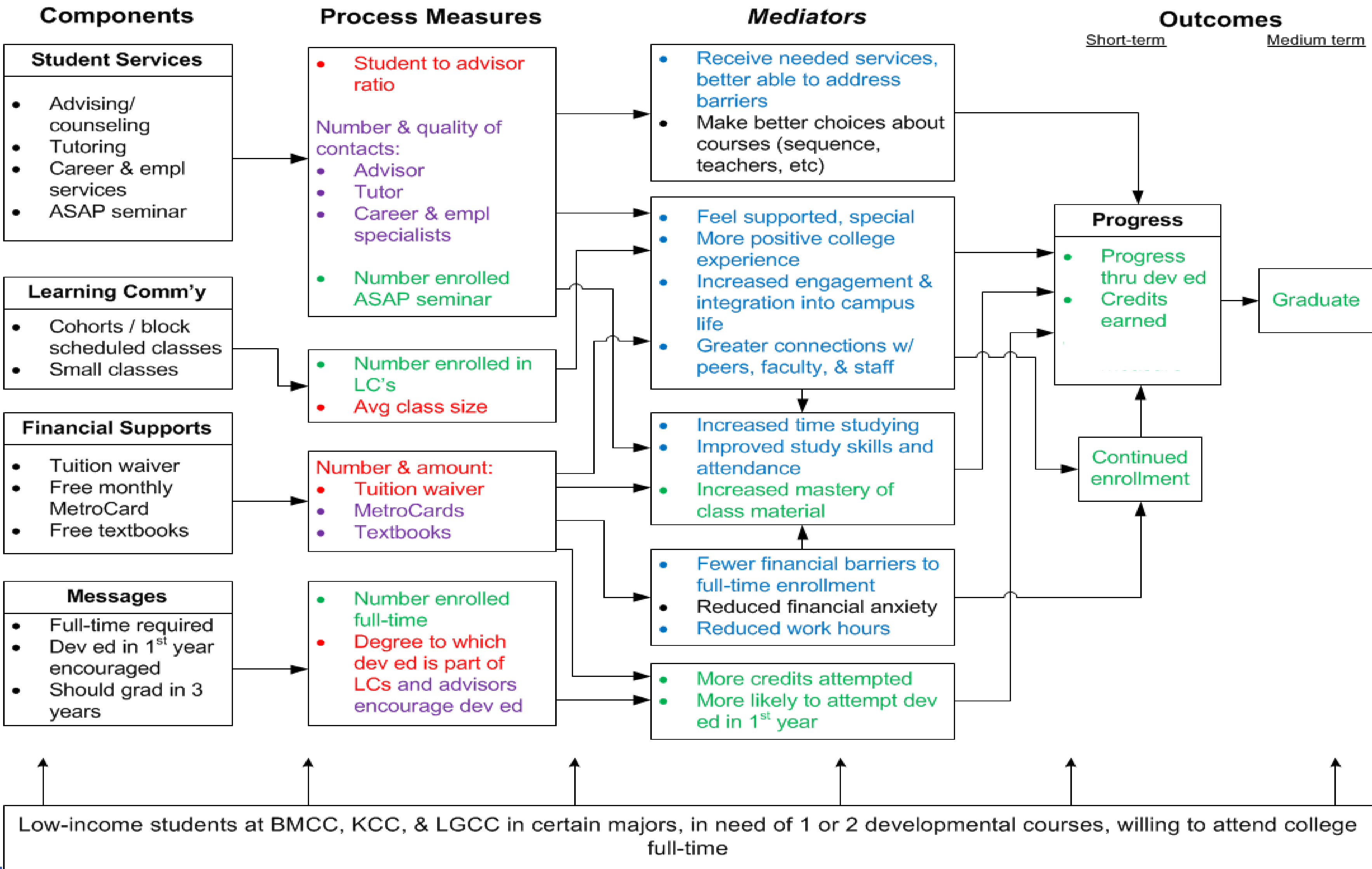
strategies, resources, or preparation that activate or support the direct components in the study and that are specified as part of the intervention model

Exhibit 4. Template for describing and collecting information for each planned intervention component

Component	Direct (D) or Support (S)	Dimension				Data sources for intervention group				
		Content	Quantity	Mode	Quality	Docu- ments	Logs: Teacher	Obser- vations	Logs: Coach	Inter- view: Teacher
Small-group student sessions in classroom	D	Modules of math concepts and strategies, with progression over modules	Weekly module: 5-minute introduction and modeling with 20-minute student activities	In-person groups of 3-4 students	Teacher-student interaction	✓	✓	✓		✓
List any additional direct components and their dimensions.										
Coaching sessions for teachers	S	Coaching session to review previous month's activities, preview next steps, answer questions, provide technical assistance	Monthly 1-hour meeting	In-person or virtual, one-on-one meetings; follow-up by email as needed	Coach-teacher interaction	✓	✓	✓	✓	✓

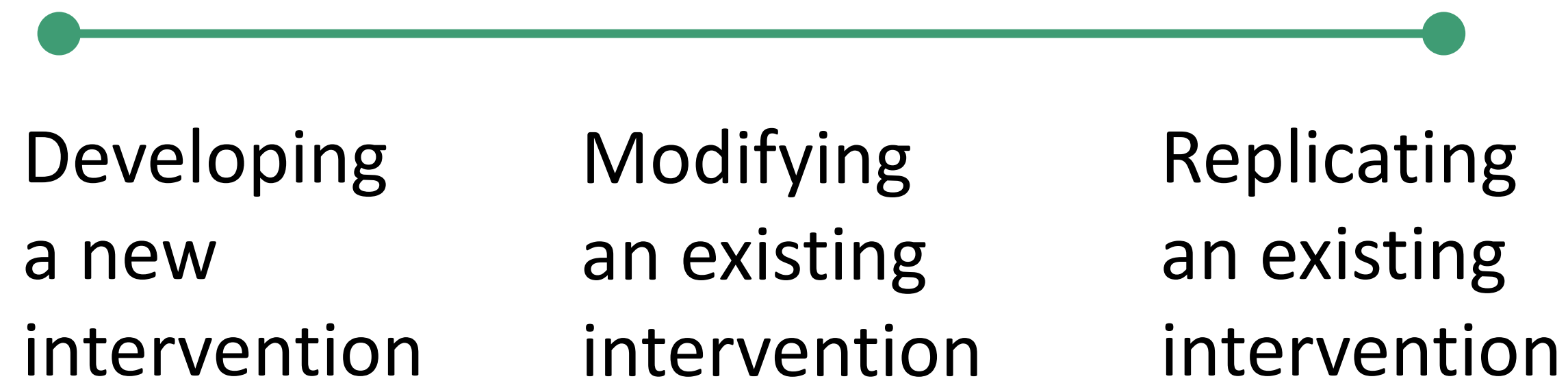
Data Source Legend:
 Green = Transcript
 Blue = Survey
 Red = Field Research
 Purple = Survey & Field Research
 Black = Unmeasured

ASAP Analysis Plan
Figure 1
Logic Model



Specifying the *planned* intervention components...

Some additional considerations



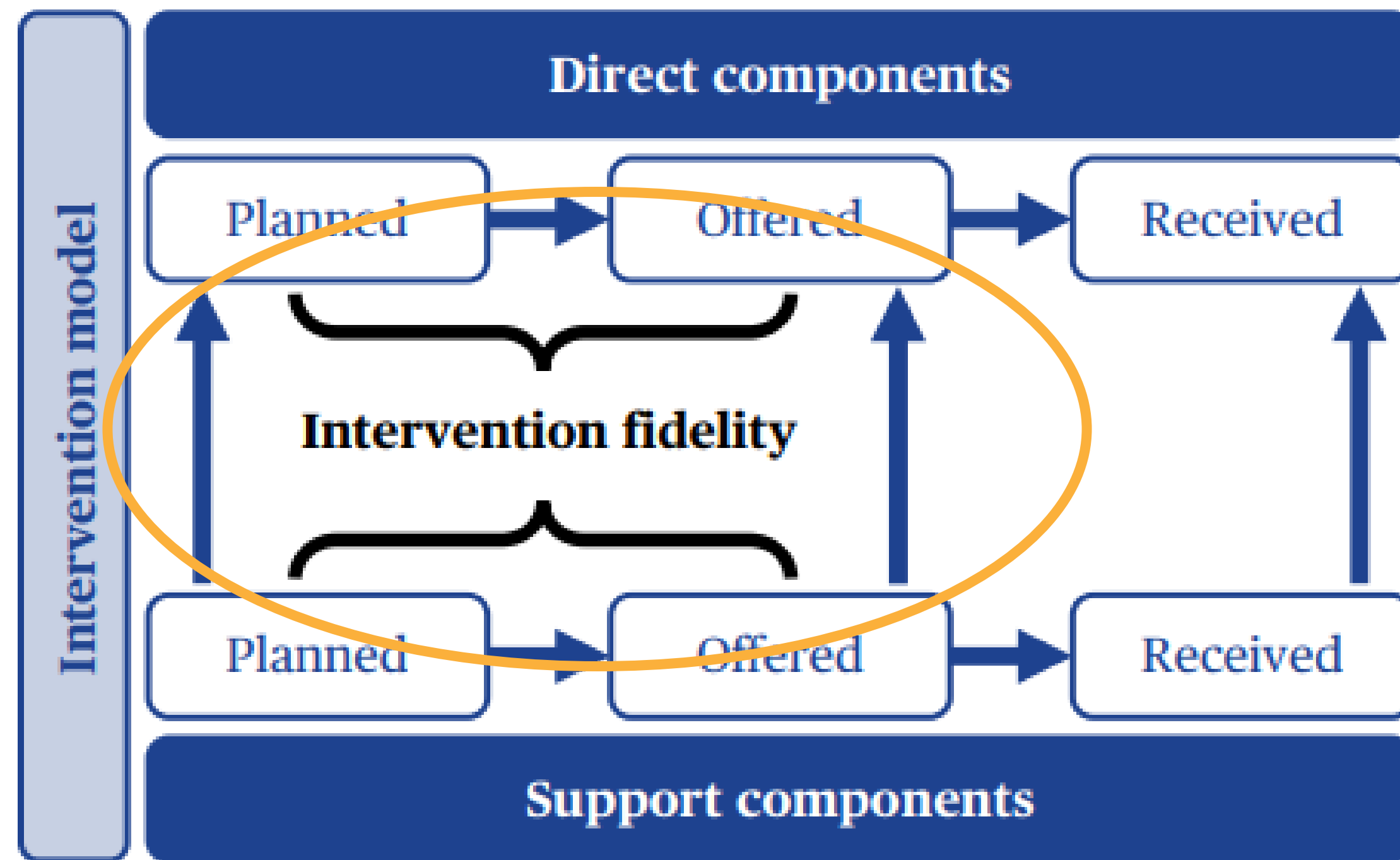
Box 2. Examples of adaptations (also called planned modifications)

Implementers, funders, developers, or your study team may suggest or require adaptations to the intervention's direct components or support components. These might look like the following:

- Adding or removing components
- Reordering components or segments
- Shortening or extending the time spent on any one component
- Breaking up content over multiple sessions
- Modifying curricular content to be more culturally appropriate
- Repeating components of the intervention, such as training or curricular content

For additional examples of possible adaptations, see Figure 1 in Wiltsey Stirman et al. (2019).

Fidelity



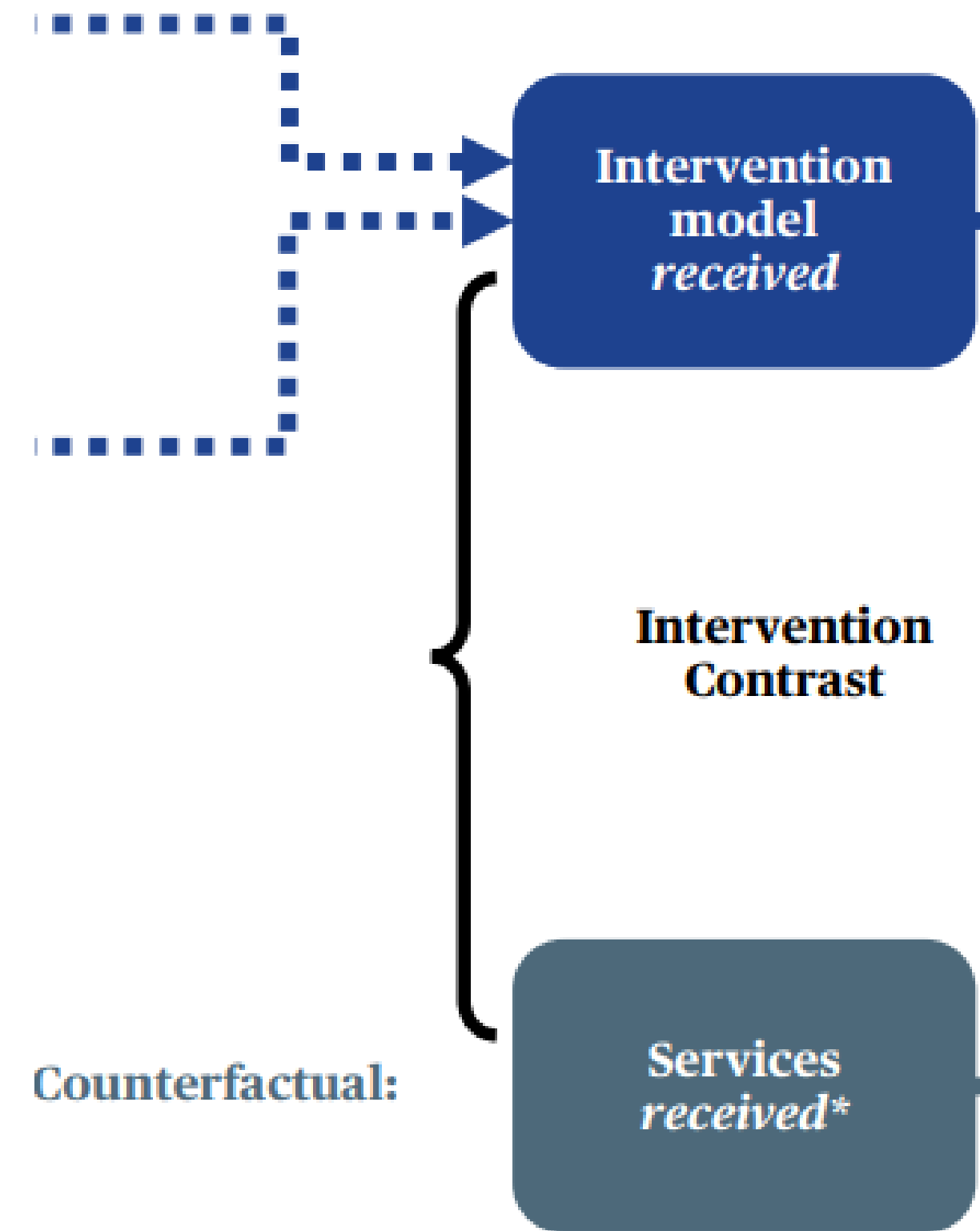
Assessing fidelity/integrity: It can be murky

- Fidelity: extent to which the intervention *offered* corresponds with the intervention *planned*
- But...there's little consensus on
 - What to call it and how precisely to define it
 - How to measure it
 - What constitutes “adequate” fidelity for any specific component or set of components
- So...
 - Be guided by the core idea: “extent to which the intervention offered...
 - Assume nothing and ask lots of questions....

Some key questions to keep in mind about fidelity *throughout a study*

1. What do key interest holders mean when they refer to fidelity?
2. Is the intervention model new? a replication? an adaptation?
3. To what extent are intervention components specified? (manualized? intentionally flexible in some or all dimensions?)
4. For multi-site studies, how do the components differ across sites?
5. What kinds of unplanned modifications during the study are possible? probable? (see Wiltsey Stirman, Baumann, and Miller 2019)
6. How might intervention fidelity vary over the timeframe of the study? Why?
7. What are *priority* components (or dimensions) for measuring fidelity?
8. What's the minimum/essential amount of intervention components that would represent (in)sufficient implementation for a “fair test”?
9. Other....

Intervention Contrast



Intervention contrast example: CUNY ASAP

Students' First-Year Experiences: Career Services Three-Year Impacts Report

Outcome	Sample Size	Program Group	Control Group	Difference	P-Value
Ever met with career or employment services staff (%)	736	79.5	28.8	50.6 ***	0.0000
Average number of times spoke with career or employment services staff					
First semester	719	4.9	1.0	3.8 ***	0.0000
Second semester	721	4.1	0.6	3.5 ***	0.0000
Survey sample size	742	384	358		

SOURCE: MDRC calculations from the MDRC student survey.

Intervention contrast example: CUNY ASAP

<u>Student Services</u>	
• Advising: Student-to-adviser ratio between 60:1 and 80:1; 95 percent of students met with an adviser during first year and students met with an adviser an average of 38 times in that period • Career services: 80 percent of students met with career and employment services staff during first year and students met with such staff an average of 9 times in that period • Tutoring: 74 percent of students received tutoring outside of class during first year and students met with a tutor an average of 24 times in that period	• Advising: Student-to-adviser ratio between 600:1 and 1,500:1; 80 percent of students met with an adviser during first year and students met with an adviser an average of 6 times in that period • Career services: 29 percent of students met with career and employment services staff during first year and students met with such staff an average of 2 times in that period • Tutoring: 39 percent of students received tutoring outside of class during first year and students met with a tutor an average of 7 times in that period

Reminder: Dimensions for *each* component...Consider corresponding relevant dimensions for control group

Dimension	Question of interest	Examples
Content	What is the component or service?	Course topics, messaging, \$, advising
Quantity	How much of the component is provided? <i>Frequency:</i> How often? <i>Time:</i> How long is each encounter? <i>Duration:</i> Over what time period? <i>Reach:</i> What % target pop reached?	•1x/week •30 minutes •2 semesters •80%
Mode	How, when, and by whom is the component provided?	In person, Tuesday afternoons, trained paid adviser
Quality	How well is the component delivered? (What would constitute “high quality” for this component?)	Tailoring to student needs, engagement, ?

Use the intervention logic model *and* knowledge of control group settings to prioritize aspects of intervention contrast that your study will examine

- Logic model provides comprehensive listing of intervention components
- Logic model shows hypothesized linkages between components, proximal outcomes, and distal outcomes
 - Are particular outcomes likely to be activated by particular components?
 - What are implications of strong or weak intervention *contrast* for chances of seeing *impacts* in those outcomes?
 - What are your study's *priority* components for measuring contrasts? On what dimensions?
- Do everything feasible to understand and document “Business as Usual” (BAU) control conditions

Intervention contrast – some key considerations

- It is possible to use the same data source(s) for both intervention and control groups?
- At what point(s) during the study will you measure intervention contrast?
- On which components, for which dimensions (content, quantity, mode, quality) is contrast *possible*?...is contrast *likely*?
- For which components, for which dimensions, are you *not able* to measure contrast? How central are those to your logic model?
- How can/should you use intervention fidelity measures for the control group?
- What would represent *insufficient* contrast between program and counterfactual services for specific intervention components?
 - Both quantitative and qualitative inquiry can be informative
 - Very risky to rely on (or hope for) the *combination* of weak intervention contrast on multiple core components or dimensions to produce impacts on target outcomes

Context Features



Table 2. Study sample and context

Key Questions	Description
What is the population of interest?	□ The population to which the study findings should apply in other contexts. □ Steps taken to increase generalizability of findings, either via sampling or analysis.
Who participated in the study?	□ Student grades, and ages for early childhood and adult learners and learners with disabilities. □ Student race, including American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, and other. Some students may identify as belonging to multiple racial groups. □ Student ethnicity, as Hispanic/Latino or not Hispanic/Latino. □ Student sex, gender, or gender identity. □ Students with an individualized education program, individualized family service plan, or 504 plan, and disability type—autism spectrum disorder, emotional disturbance, intellectual disability, developmental delay, or other specific disability. □ Student English learner status—current, former, ever, never—and other English learner characteristics—language spoken at home, newcomer, long-term, English proficiency level. □ Student socio-economic status—free or reduced-price lunch, Pell Grant eligible, other. □ Other student characteristics—homeless, migrant, foster. □ Student achievement on a standardized measure before the start of the intervention—below grade level, on grade level, above grade level. □ Educator characteristics—years of experience and credentials.
Where did the study occur?	□ Country or state(s), if in the United States. □ Urban, rural, suburban, or town setting. □ Number of schools, postsecondary institutions, or educational sites. □ School type—charter, parochial, public, or private—and format—in-person, online, before or after school. □ Postsecondary institution type—two-year, four-year, public, private. □ Other educational site—center, home-based. □ Classroom type, including general or inclusion, self-contained special education, and designated English language development. □ Other school, institution, or site characteristics—enrollment, Title I status, magnet, student characteristics.

Example characteristics of external environment (outer setting)

- Policies or regulations (federal, state, district, system)
- Financing structures (federal, state, district, system)
- Economic conditions, labor market, competition
- Geographic aspects
- Historical or sociocultural aspects
- Political aspects
- Shocks (e.g., pandemic)

Example characteristics of organization/system (inner setting)

- Mission/values
- Decision structure
- Resources (physical resources, human capital)
- Leader vision/characteristics/behaviors
- Organizational culture or climate
- Presence/actions of unions
- Readiness for change (structural and psychological)
- Student/client population (e.g., number of students, aggregate student characteristics such as race, ethnicity, socio-economic characteristics, absenteeism, mobility)
- Other initiatives being implemented at the same time as the intervention

Example characteristics of individuals (students, teachers, staff, leaders)

- Demographic characteristics
- Educational background
- Professional background and work experiences
- Attitudes and beliefs
- Social-emotional characteristics (e.g., burnout, depression, anxiety)
- Psychological factors (e.g., readiness for change)

Which context features should you measure and report in your study? Some considerations

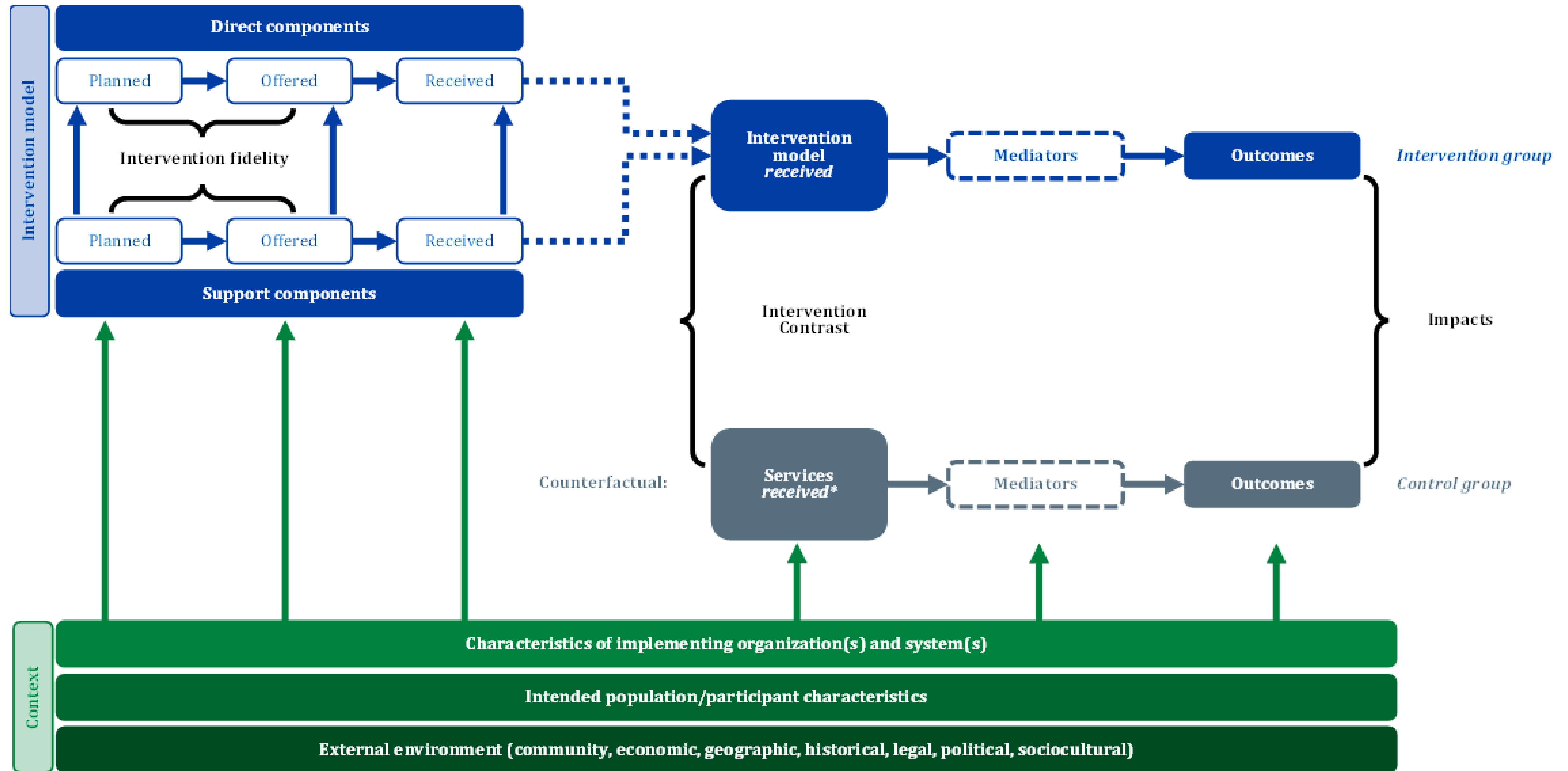
- What are the requirements or recommendations from funders or other interest holders?
- What context features does relevant prior research point to?
- What context features are most relevant for the current study?
 - Likely to vary within or across settings or time?
 - Likely connected to intervention contrast?
- What information is realistic/feasible to collect during the study?
 - Burden on study participants and intervention implementers
 - Financial cost
 - Data quality
 - Time frame of the study and data availability

Resources:

<https://www.youtube.com/watch?v=EazeFfFfXr4>

[Prioritizing and Selecting Context Features in Education Impact Studies: A 4R Lenses Approach](#)

Exhibit 1. A conceptual framework for linking the intervention model, contrast, and impacts



Have a great Institute experience!

Please contact me with any questions and feedback

Carolyn.Hill@mdrc.org