STATE AGENCY AND SYSTEM REPORTING TO AND USE OF POSTSECONDARY DATA PARTNERSHIP DATA FIELDS

Carrie Klein and Christina Whitfield

ACKNOWLEDGMENTS

This report is based on research funded in part by the Bill & Melinda Gates Foundation. The results, recommendations, and conclusions contained within are those of the author(s) and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

INTRODUCTION

The Postsecondary Data Partnership (PDP) was developed by the National Student Clearinghouse to help higher education institutions better use their student-level data. The PDP is a nationwide tool that aims to improve the ability of institutions, states, and systems to collect, access, and report their data with greater ease. Through data fields that are tied to common key performance indicators and metrics, an analysis ready file (ARF) that allows institutions to benchmark data against other institutions or states, and data dashboards that provide digestible visualizations, the PDP's goal is to support and improve postsecondary outcomes.

While the focus on the PDP has been at the institutional level, a number of states and postsecondary systems have engaged the PDP. State-level participation has been integral to PDP development since its inception. As of this report, 17 state agencies or systems participate in the PDP, with four additional in the process of joining. These entities are composed of state postsecondary system agencies, boards of regents, and community and technical college systems. These state agencies and systems either submit their statewide or system-wide data on behalf of their state's institutions or use the data submitted to the PDP by their institutions. The former approach is often made by state agencies and systems with robust internal data systems and centralized reporting processes, while the latter approach is more often used by state agencies or systems with less robust data infrastructure and decentralized reporting. State agencies and systems also share PDP resources and communications with their institutional members to support institutional-level participation. As a result, state entities have been interested in understanding how the PDP can be used to support their policymaking and decision-making priorities.

Through convenings, reporting, and technical assistance, the State Higher Education Executive Officers Association (SHEEO) has supported our state agency and system members and the broader postsecondary data community's state- and institutional-level participation in the PDP. Through our

work, we have developed a toolkit with resources, including data crosswalks and submission guides, blog posts, and white papers. SHEEO has also partnered with the Association of Public and Land-grant Universities (APLU), the American Association of State Colleges and Universities (AASCU), and Jobs for the Future (JFF) to provide technical assistance to their member institutions who are submitting data to the PDP. The goal of this work has been to help reduce state and system reporting burdens while also promoting connections, communications, and best practices across various state-level data systems and communities.

This paper is a part of that work. The purpose of this study is to understand state agency and system reporting patterns to PDP data fields and their associated use and experience of PDP's ARF and dashboards. The following research questions guided this inquiry:

1. What trends/patterns exist in state-level PDP data field reporting?
2. What aspects of the PDP field reporting, dashboard, and ARF have been useful to states agencies and systems?
3. What aspects of the PDP field reporting, dashboard, and ARF create challenges for states agencies and systems?
4. What do state-level participants value in the PDP fields, dashboards, and ARF?

This approach allowed us to better understand state agency and system experiences with the PDP, through review of state agency and system submission data to which we had access and by conducting a survey and interviews with state agency and system participants.

Among our findings are that state agencies and systems are able to report to a majority of the PDP fields with relative ease. Limits in reporting are often due to a lack of data from the state or system, a lack of alignment in data definitions, or challenges with matching data across disparate systems. Overall, state agencies and systems see the value of the PDP reporting fields, as those fields connect well with state policy priorities. Further, the state agencies and systems we spoke with as a part of this

project have shared PDP resources and communications with their institutional members to support institutional PDP use in their states.

However, the value of the PDP to a state agency and system is influenced by the maturity of their internal data systems. State agencies and systems with more robust and developed data systems (the bulk of those existing in the state agencies and systems represented in this study) are less likely to find utility in the PDP, as they do similar collection, analysis, and reporting in-house. However, for state agencies and systems without a robust data system (for instance, those lacking data warehouses), access to the PDP and its ARF was viewed with great value. State agency and system resource and capacity contexts, also aligned with level of data maturity, added another layer of complexity in understanding PDP reporting and participation.

METHOD

To understand state agency and system perspectives on reporting to PDP fields, we employed a mixed-methods approach to data collection and analysis. In August 2021, we began by running a descriptive analysis on the ARF data for states in the PDP system for which we have access to determine data field reporting patterns. To provide context for the reporting trends that our review of the ARF highlighted, in September 2021, we met with current state-level PDP participants to ask about their experiences with submitting data to and using the PDP. That check-in meeting informed the data collection process, which included developing and distributing a survey and conducting individual interviews. We analyzed the data from the descriptive file, survey results, and interview transcripts, which informed the themes discussed in the Results and Recommendations section. More information on participants and on data collection and analysis are listed below.

Participants

We collected data from state-level participants (either current or former), those contemplating participation in the PDP, and those who had considered and declined participation in the project.

Participants' roles ranged from data analysts who regularly work directly with PDP data to vice chancellors who oversee the strategic participation in PDP initiatives in their states. Outreach to participants, via email, included the current state-level PDP participants who had participated in the initial check-in meeting (13 in all) and members of SHEEO's Research and Data Group (82 in all). When data were collected for this report, 16 state agencies and systems had contracted with the PDP (see *Table 1*, Appendix B). State agencies and systems participating in this study were at various stages of participation. Some have been involved in the PDP since its inception, and others had just recently submitted data. Through our third-party affiliation with the PDP, SHEEO has access to seven of the 16 state agency and system level participants' data.

Analysis Ready File (ARF)

Data from the seven participating state agencies and systems (including 2,128,581 student-level records) to which SHEEO has access comprise the ARF we used to determine data field reporting patterns. In late August 2021, we ran a descriptive analysis of the ARF to determine which data fields were reported to most frequently and which had the highest level of missingness (see Table 2 in Appendix C). The results from this analysis are described in the Results and Recommendations section of this report. While the results were informative for illustrating reporting patterns, the data file does not indicate why these patterns exist. To better understand why state agencies and systems were reporting as they were, we used the results from the ARF to inform survey and interview questions.

Survey

We deployed the survey in early October 2021 to all state-level PDP participants and members of SHEEO's Research and Data Group. The survey remained open through mid-November 2021 (see Appendix A). Closed and open-ended questions focused on state-level participants' experience with submitting to the various PDP fields; their use of PDP dashboards and analysis readiness files; and if and how the various data fields tracked with state policy priorities. We received nine responses to the

survey. One of these responses was by a participant in the original check-in meeting. This participant agreed, along with six other check-in meeting participants, to participate in an interview to dive more deeply into the insights the survey provided.

Interviews

We conducted semi-structured interviews with seven of the 13 check-in meeting participants from October through mid-November of 2021. Each interview lasted from 30-45 minutes. We asked questions that focused on which fields PDP state agencies and systems reported to, experiences associated with that reporting, and the alignment of PDP fields to state policy priorities. We also asked about the value and use of the ARF and dashboard components of the PDP to state work and any other perceptions or experiences related to data submission or use that participants wanted to share.

Data Analysis

We reviewed the descriptives from the ARF and close-ended survey answers and used open, hand coding to analyze data in the open-ended survey answers and interview transcripts. Through that review and analysis, we constructed two major themes associated with state-level PDP reporting and use. These themes focused on 1) the reportability of various fields and how the decision to report to a field is influenced by participants' experiences, ability, priorities, and trust in the data, and 2) the variability of perceived value of the PDP as viewed through its connection to state policy priorities, data maturity, and in response to resource and capacity constraints. These themes, along with an explanation of the ARF descriptives, are described in detail in the Results and Recommendations section of this report.

RESULTS AND RECOMMENDATIONS

Inquiry into the reporting and use of PDP data indicates disparities in the frequency with which state-level PDP participants report to various fields. Participants noted that once data were aligned with PDP standards, submission to fields was relatively straightforward. The gaps in field reporting generally

came down to an absence of available data, a decision not to report available data, or a lack of confidence in data definitions and associated benchmarking. All participants noted a strong synergy between state policy priorities and key performance indicators and the PDP data fields. However, participants conveyed varying degrees of value in PDP participation, which was influenced by the status of their data systems. Participants from states with more mature data systems and robust data collection efforts found less overall utility in PDP reporting and participation, while state participants with less mature systems and without centralized data collection mechanisms found great value in PDP reporting and use, especially use of the ARF. Finally, participants noted various factors informing their PDP participation, many of which are related to resource and capacity constraints, including institutional reporting burdens. The results, below, are paired with solution-oriented recommendations for improving alignment between state agency and system needs and perspectives with PDP tool capabilities and support.

PDP Field Reporting Overview via the Analysis Ready File (ARF)

Over two million student records are represented in the combined state-level ARF to which SHEEO has access. State-level PDP participants are able to submit to 56 fields (not including financial aid data fields, which are submitted via a different process), with SHEEO having access to 48 of those standard data fields (less student identification fields).[1] These fields include enrollment, attendance, cohort, retention, completion, persistence, time-to-degree, gateway course, developmental math and English course participation, dual enrollment, and demographic metrics (see Table 2). The state-level participants represented in the ARF universally submit enrollment, cohort, attendance, persistence, and retention metrics. Participants reported to fields associated with earned and attempted credits at

[1] Note: A number of data fields exist beyond the list presented in this table; however, these data fields are those that SHEEO has access to via state-level ARF data that have been shared with us by state agencies and systems. For a complete list of data fields and descriptions of each field, see: https://studentclearinghouse.info/onestop/portfolio-item/postsecondary-data-partnership-submission-guide/

98.91%, reported GPA for year one at 98.15%, and reported math and English placement data at 81.64% and 81.5%, respectively.

Regarding student demographics, student race and ethnicity were reported with great consistency (98.34% and 97.49%, respectively); however, student gender reporting was less robust at 89.7%, Pell eligibility in the first year at 82.21%, and, notably, student first-generation status reported at only 28.27%. Additional fields with missing data at more than 70% include the time-to-credential field, fields associated with gateway course status, attempts, completions, and grades; developmental math and English course attempts and completion; and dual and summer enrollment. The fields with the greatest percentage of non-reporting were dual and summer enrollment (only 1.6% reported), math gateway grades (only 4.74% reported), and English gateway grades (only 6.11% reported). For additional details on field reporting by data element, see *Table 2* in Appendix C.

PDP Reporting Experiences: Data Availability, Alignment, and Visualization Informs Submission and Use

Participants noted that most PDP data fields are easily reportable to, once data matching and definitions are aligned. There are some fields where state agencies and systems are unable to share data, historically have not collected the data that PDP is asking for, or no statewide standardization or definitions exist. This context limits the ability of states to report to these fields. Because of varied data definitions or unavailable data, participants noted that PDP data from these fields are less useful for them (even when acknowledging their potential usefulness). Missing or misaligned data influenced participant trust in using data from these fields for benchmarking purposes. Further influencing PDP use were the PDP dashboards associated with data field reporting. Participants noted the complexity of the dashboards, which impacted load time. The positioning of these dashboards toward education professionals (rather than agency or system stakeholders) means that in their current form they are not useful for their work.

Data Fields: Reporting and Use Challenges and Recommendations

Participants across states and systems noted that, despite the general ease of submitting data to the PDP, there are a few fields that created particular challenges. Challenges were primarily associated with workload, matching data to align across states with different contexts and definitions, and the availability of what data states have access to, collect, and are willing to share. Among the most referenced data fields were address, Classification of Instructional Programs (CIP) codes, gateway and developmental course, student status (e.g., first-year freshman) and demographic (e.g., first-generation, gender, and race and ethnicity fields), and financial aid fields.

Address Field: Among the most cumbersome data fields for participants to report to is the address field. This field created significant consternation for state agency and system participants. A participant from a small state system explained that the address field was cumbersome to report to because "the address had a limit of 20 characters, so it required clean up." Multiple participants noted the challenges of reporting to the address field and did not understand why the PDP needed student addresses. For example, as a participant noted, "I don't know what they're using address for, but it's like going through 280k addresses that might have errors in them, you know, is a whale." Participants indicated that this field, in particular, required a heavy lift by staff for limited payoff for their state or system.

> *Recommendation: Although the address field is essential to PDP's student identifier matching algorithm (it is used to confirm student identification across institutions and states), state agency and system participants are unaware of why they are being asked for this data. To increase awareness, the PDP could consider providing additional detail in their data submission guide and support communications to explain to submitters why this data is needed. Further, if there is a way to allow for submission of address data beyond the 20-character limit, that might help lessen the burden on institutions.*

CIP Codes and Gateway Course Fields: In addition to the work required to input address information, misaligned data definitions and state and institutional context also created challenges to data reporting. Participants noted the CIP codes and Gateway Course data were among the most challenging to report because of varied definitions based on state and institutional differences. This meant aligning their state -or institutional-level data to PDP data was a difficult process.

A survey respondent noted that assigning CIP codes to courses took the majority of their submission time, and state and PDP definitions were often not aligned. Another survey respondent concurred, noting that their "campuses indicate assigning CIPs to specific courses has been awkward/cumbersome." Part of what makes this field awkward and cumbersome is that different states and institutions may assign CIP codes differently to courses, which complicates use of that data for analysis and benchmarking. This is particularly true for courses that span programs, as this participant from a college system explained:

> [The system institutions] had to assign CIP codes to every single course, and our campuses are like, we don't know what to do, like what if it's a course across multiple programs? There was lots of confusion, but our folks eventually figured it out. I don't think they did them consistently across our systems, so that's probably going to be a situation.

Because CIP codes are assigned differently across institutions and not systemically or consistently used at the course level, the data in the PDP, while aligned, are not necessarily indicating the same information.

Like the definitional and contextual differences in CIP codes, participants noted that gateway English and math courses created similar challenges. A participant from a community college system explained:

> Gateway courses are something that really differs across institutions. We are constantly having to shift our developmental education policy [internally] as far as who is college ready and comparing and doing analysis and cutting students differently across time is tricky, because you should be coding on how things were at the time, based on standards then, so that is why we left some of those fields blank because, you know, the requirements are always changing just based on what semester we are in.

Because these changes are continuously happening and do not always align across institutions or states, this participant indicated that they do not use gateway course fields in the PDP. A participant of a centralized state system indicated the same, noting that both PDP definitions and their state context limit the usefulness of the gateway fields. "Our community colleges allow students to take classes at any campus without any transfer, then those numbers [in the PDP] are not useful for us, because it's way under the constant percentage of students that are taking English and math in the first year because they can take it at another campus." Participants indicated that due to variability and availability of gateway course data, even if they can collect those data, they are often difficult to align with PDP fields. They noted the same for data associated with dual enrollment and developmental courses.

> *Recommendation: Despite the acknowledgment by members that the PDP has done a good job of trying to align data across contextual differences, the varied abilities to collect these data and disparate definitions for these fields limit their effectiveness for use by state agencies and systems. To improve their utility to state agencies and systems, recommended are intrastate and interstate convenings of PDP agency, system, and institutional participants to inform ongoing definition development for these fields and communicate ways to use these data to create insight, despite contextual differences.*

Student Identity and Demographics Fields: Definitional and contextual differences extend to a number of other fields in the PDP. For example, a participant from a small state system with a young

data system explained that their definition of first-generation students differs from the PDP's: "We have to match the PDP . . . [definition of] first generation in our systems – [we have] a different value compared to the PDP." Similarly, a state system with a robust data system noted the challenges that state context and definitions play in aligning state data to PDP standards. As they explained in the use of the student ethnicity fields, "This state has a high degree of ethnically diverse students and wants to be able to further disaggregate the data, so the PDP reporting doesn't help us the way we need." Because the PDP's race and ethnicity fields do not match the further disaggregation of this state's student ethnicity data, it is not useful to the agency's policymaking and decision-making.

> *Recommendation: Allowing for further disaggregation of student identity data or better aligned definitions of data would allow for more contextualized use of these data by state agency and system participants. Given the growing need of state agencies and systems to understand and support nuanced student contexts in relation to state completion and success initiatives, allowing for greater disaggregation and contextualization of data (especially race and ethnicity data) should be a priority.*[2]

Financial Aid Fields: Beyond availability and matching issues, legal concerns limited state and system reporting to PDP's financial aid fields. All participants noted that they had decided not to submit financial aid data because of privacy concerns. As a participant explains, "We chose not to report the financial aid just because [our counsel] were questioning whether we should be sharing that data with anybody." Concerns about privacy violations in relation to financial aid data permeated the interviews with participants and limited the usefulness of this field to their work. This is unsurprising, as, "despite

[2] In January 2022, the PDP announced at a Postsecondary Data Collaborative meeting that it would further disaggregate gender data and would consider further disaggregation of race and ethnicity data.

their importance to informing college completion and student success outcomes, financial aid data are often employed inconsistently in higher education due to a lack of federal policy governing use."[3]

> *Recommendation: The PDP continues to convey to state and institutional participants that financial aid data are secure within the PDP system and has seen increased participation in financial aid submissions. Until there is more comprehensible and consistent federal guidance, the PDP should continue to promote the importance of secure financial aid data sharing and work with state agencies and systems to advocate for this reporting from their institutions.*

Consequences of Misaligned or Mismatched Data: More than just not useful, misaligned data and definitions can also result in incorrect data. For example, a participant from a large, centralized state system with a robust data system noted that PDP data definitions differ from their state's definitions (which are based on IPEDS reporting standards):

> In our state, a student could technically be a first-time freshman in Fall 16 and then be a first-time again in Fall 20. We also have what are called "summer returning fall first time freshman" and so the student comes out of high school, takes some summer courses, but we still consider them a first-time freshman in the fall. Because the [National Student] Clearinghouse does the data validation for who's included in the cohort, it makes it impossible to include those students in your first-time freshman cohort. And so, you know the first couple of times I did the submission process, I just took all of our first-time freshman and put them in the cohort files and got an email back which said, "Hey, you know, we found that you have 250 duplicates." And, so I had to come up with a process where I took out that cohort and then hit it again against historical submission and then took out anybody that was already in a previous file.

[3] SHEEO (2019, Sept. 23). Making sense of federal financial aid data use policy for research and evaluation: visioning a state role. State Higher Education Executive Officers Association.

In addition to creating additional work for the state or system participant, misaligned definitions result in misaligned data, reporting, and benchmarking. This impacts trust in the PDP data fields to inform policymaking and decision-making.

As a result, the participant noted that "we have historically just not reported those data elements because it's not clear." This causes institutions to question the data, as the participant noted, "When our schools go into the dashboards or something like that and they say look, we've tried really hard to match this cohort and we can't do it, or we don't understand why the completion rate that's been reported by the PDPs is so much higher than what we report internally." Further, the responded noted that when they encounter these issues, they are "less likely to use the information." If states and their institutions cannot trust data because of misalignment, they will not use the data, ultimately limiting the value of PDP reporting and benchmarking.

Recommendation: Where possible, PDP fields should align with pre-established federal standards to help alleviate the burden on state agencies and systems to report data. Doing so would help standardize reporting and bolster data, benchmarking, and dashboard use.

Complex Visualization Influence Use and Integration:

Further impacting consistent use of the data reported to PDP data fields by state agency and system participants are the data dashboards. Although participants noted that cross-state benchmarking, particularly for enrollment and completion metrics, is a priority for them, they do not use the dashboards to help with that benchmarking. Problems with loading and the complexity of the visualizations and how they are communicated limited their usefulness to state participants. As a result, the participants rarely use the dashboards and associated benchmarking to inform their work. Instead,

participants rely on their own, internal development of data visualizations, which they felt were better aligned with their needs and better informed by their state context.[4]

Capacity and Complexity: The ability of state agencies and systems with robust data systems to create and use internal dashboards as a preference to PDP dashboard use permeated the survey responses and interviews. One participant from a state that was an early PDP adopter but has since stopped participating in the PDP noted that "we had built stuff internally that met the needs for what we were doing. We have dashboards that we've created for our campuses around retention and graduation…We produce some state reports of our campuses…but we just haven't used the PDP dashboards." Because this participant's agency can do bespoke dashboard work in-house, they do not see a value in using the PDP's dashboards.

Other participants noted that, regardless of their in-house ability to create dashboards, the complexity of PDP dashboards limits their use. For example, a participant from a large state system noted that, "Those dashboards have so many filters and in some ways, they are pretty non-traditional as compared to what people would typically look at for reporting that I think it's just a little bit overwhelming…and it also takes the dashboards a really long time to load." For example, the participant noted that they often hear from institutions that it takes more than 30 minutes for their dashboards to load. This matters for use, because, as the participant noted, when "the president asked the question two hours ago and they still haven't gotten the answer [because the dashboard will not load]," that is a problem. For this participant, multiple filters, load time, and a mismatch in practical application of dashboard data make the PDP dashboards ineffective for informing work.

Recommendation: To improve uptake of the dashboards, reducing the number of filters might help reduce load times. Checking in with participants to determine which filters are most

[4] In December 2021, the PDP addressed the dashboard loading times, which have subsequently improved.

> *important might help with that determination. Querying states and systems (whose needs might differ from institutions) could help develop a more streamlined and useful set of filters and metrics to inform faster-loading dashboards. One participant noted if PDP employed an Online Analytical Processing (OLAP) cube[5] approach to allow for tailored filtering and structuring of data and visualizations, it might help improve dashboard utility and uptake.*

Accessibility and Understandability: Multiple participants noted that they would rather create their own visualizations because, as one participant explains, "I think how we do visualizations is much easier for people to interact with from an end user perspective and they're totally accessible, which is a big difference between the [internal visualizations and PDP dashboards], ... and knowing we are already fulfilling that need within our state makes the need to use them less." While this participant and the institutions in his state do not use the dashboards, they do think PDP dashboards can be "a really big benefit for those states and institutions that don't have those tools available to them."

But beyond loading and visualization challenges is concern about how the dashboards are positioned. A participant explained:

> To be honest, I don't ever use the dashboards. One of our limitations is that anytime we show information to our board, it has to be public. Even on our board, who, you know, they're non-education professionals, most of them have not worked in education in their careers, so we have to be very thoughtful about how we present data to them…We are very thoughtful about what we do and how we articulate it, so we don't use terms like "matriculation," we use terms like "started college." So, the PDP dashboards I think are designed for educators in mind.

[5] An OLAP cube approach to data analysis allows for multi-dimensional use of data from varied perspectives and allows more nimble filtering and grouping of data by users. For more information see: https://docs.microsoft.com/en-us/system-center/scsm/olap-cubes-overview?view=sc-sm-2019

Because PDP dashboards use language from reporting fields that are particular to education professionals and internally positioned, they are not always useful in communicating information to broader state and system stakeholders. This limitation, in addition to their complexity and load time diminishes their value for state and system participants.

> *Recommendation: As the PDP develops dashboards intended not just for education professionals (as PDP currently indicates) but for a broader set of stakeholders, it will be important to include the groups to whom state agencies and systems report. State agencies and systems can help inform appropriate language to use when positioning dashboards for external stakeholders.*

Varied Value of PDP Reporting and Participation

All participants noted a clear connection between PDP data fields and their state policy priorities. From their perspective, this was among the strongest attributes of the PDP's data fields. They also valued the potential for benchmarking that PDP participation provided. However, perceived value in PDP data use and participation varied. States without centralized data or more robust systems saw great value in PDP participation and especially in the ARF.

Conversely, states with more modernized and developed data systems indicated that use of the PDP was not of particular value to their work processes or reporting because they already have similar or, in their view, more appropriate and applicable data and reporting capacity in-house. Participants also communicated concerns about the cost associated with PDP participation. For more established or robust data systems in states, the cost-benefit analysis resulted in the decision not to continue or to sunset participation. However, for states with nascent systems, they felt the cost was well worth the ability to gain access to the data in the ARF and to build their capacity to do data work, as an alternative to the cost of establishing a data warehouse and data system in their state.

Clear Connections Exists Between State Policy and PDP Data Fields

A clear value of the PDP to all participants was the alignment between PDP data fields and the policy priorities of their states. One participant from a centralized state system with a robust system noted that the PDP data fields "are actually very well aligned…They are all very policy salient for us. They're exactly what we've been working on to improve student success in the past few years." Similarly, a participant from a college system with no centralized data system noted that "almost all of the KPIs we identified were included in the PDP…We are shifting our performance indicators to be around five areas—enrollment, progression, completion, affordability, and post-graduate outcomes—and PDP gets us those first three." Respondents clearly see the connection between reporting to PDP data fields and how those fields can be used to inform policymaking and decision-making work in their state.

PDP's benchmarking capabilities were viewed as a positive by all participants for providing insight and connections between their and other states' data and policy priorities. As one participant from a state with a mature data system explained:

> Benchmarking to other state systems or community colleges across the country, that's probably the most valuable for us. We feel comfortable with the resources that we have for our ability to build dashboards and those types of things, and so for us, [the PDP] is more about what it can provide us about other institutions outside of the state.

Although they expressed value in the PDP, they also noted that "we're obviously not submitting anything that we don't already have and our institutions are the same way, and then you know, our preference is to use our internal definitions that we've set for benchmarking." As described earlier, states and systems with robust data systems prefer to use internal definitions for data fields, and this extends to their use of those fields for benchmarking–ultimately reducing the PDP's utility.

Recommendation: The PDP should continue to promote the synergy between PDP data fields and state policy priorities and the possibilities of interstate benchmarking. To add value, consider adding data fields that collect affordability, accessibility, and post-graduate outcomes data and communicate ways to connect and use PDP interstate benchmarking across the P20 data spectrum—an area of increased focus by state agencies and systems.[6]

Data Systems Maturity and Organizational Capacity Inform Value and Guide Choices

Notably, participants expressed differences regarding the value of PDP reporting and the use of the ARF and dashboards. Participants with more developed, mature data systems were less inclined to see value in the PDP for their particular state or system. Conversely, for states or systems with less developed data systems, the PDP is viewed as a valuable tool for improving data-informed policymaking and decision-making.

Data Systems Maturity: States and systems with well-developed data systems are often able to collect, analyze, and use data in ways that are similar to, and even exceed, the tools and insight PDP provides. As one survey respondent noted, "Our state SLDS provides a higher level of data than current PDP." Another survey respondent explained that their system's "SLDS and internal persistence and completion reporting goes further than the PDP." Because they are able to conduct a higher level of analysis and reporting through their internal data systems, they value the PDP less. In fact, all participants coming from states or systems with well-developed data systems indicated that they do not use the ARF or the dashboards because they already have the data from ARF files in their systems and because they can create their own tailored visualizations that better meet the needs of their stakeholders.

[6] The NSC is working to incorporate wage data into the PDP, with initial integration planned for summer 2023.

Yet, for state agencies and systems with less mature data systems, the PDP has helped them leverage their data in ways they would not have been able to otherwise. A participant noted that their state does not "have a warehouse, and [PDP participation] was the fastest way to get organized." Similarly, a participant from a smaller system noted that:

> We are new, so I always think that's important context, because we're just getting started…Our data is all decentralized and it's all currently managed on Excel spreadsheets, and it drives me crazy…We found the PDP solved a lot of the challenges that we face because we don't have a shared student information management system and we don't have a well-developed institutional research strategy in that there are only two of us. So, the PDP enabled us to kind of outsource it to a trusted vendor…It would have taken us years to get to the point where PDP currently is.

For these systems, the PDP provides access to a centralized and robust data system and tools for analysis that can help inform policymaking and decision-making. As one participant noted for them, the ARF, which includes data from the institutions in their system, "is like gold" as they have no centralized data systems to which their institutions can report.

However, as states increasingly invest in and develop their internal data systems,[7] reliance on tools like PDP may wane. For example, a participant noted that their state is currently working on their data warehouse and PDP participation acts as a helpful stopgap measure. "We are transitioning to a warehouse. Our expectation was that once the data warehouse came on board that we would transition away [from the PDP]." While the PDP provides immediate value to the system, as it develops its own data resources, that value has begun to diminish. This participant went on to note that once their warehouse is online, they will stop reporting to and using the PDP for their system.

[7] SHEEO (n.d.). Strong Foundations Survey: Explore Survey Results [website]. State Higher Education Executive Officers Association.

Recommendation: The PDP provides real value to state agencies and systems with less robust data systems. Future iterations of the PDP should focus on the needs of states and systems where PDP has the most value add. Creating a focus group or advisory council of these states and systems could not only help inform PDP development but create a community of change associated with PDP use.

Organizational Capacity: Beyond the technological levers, state participants noted that organizational capacity plays a role in decisions to report to and use the PDP. As with other aspects of state PDP interactions and perceptions, there is a divide between those state agencies and systems with less and more mature data systems. Those with less robust data systems or fewer data resources view PDP reporting and participation as capacity adding, while those states and systems with more robust data systems see PDP reporting and participation as capacity diminishing and, as a result, either burdensome or not worth the cost when they can do the same work in-house.

The resource constraints facing many of the state agencies inform their policymaking and decision-making to report to and use the PDP. For example, one state agency with a robust data system that has participated in the PDP since its inception has decided to no longer continue reporting to the PDP. “We would have continued submitting data on behalf of our state’s institutions, but we lost 10% of our budget for FY21 and we aren’t getting it back–it’s just gone. So, to say I am going to spend $80K [cost of system reporting to PDP] to send you data is less in the cards.” The cost of continued PDP participation within an increasingly resource-constrained environment was too great for this participant’s state.

Because they are able to do most intrastate and some interstate benchmarking via their data, it leads them to question the value of paying to report to the PDP. Another participant noted that given their abilities, “it becomes a question of, well, is it worth it to pay that amount just to have external to

[our state] information?" This is especially true as many states are able to benchmark their data against other states' data with other data tools.

Further, data mature participants indicated that PDP participation is unnecessarily additive for them and their institutions:

> The challenge is…do we submit less information [to the PDP] and make the tool less valuable or do we add an additional burden on the institutions to submit more data? So that's been the risk [versus] reward that we've weighted with what we're currently submitting, and, frankly, we're always kind of trending in the direction of submitting less, so that we're not pulling in additional information from the institutions unless we have additional uses for that outside of the PDP. [The PDP] just doesn't' have enough benefit to us to go through that process.

Without PDP participation, this agency has the capacity to collect and conduct analysis on data from its institutions. Adding a request to institutions to report data fields to the PDP becomes another request and, as a result, reduces institutional and agency capacity.

Yet for state agencies and systems with less robust data systems, PDP participation is worth the cost. The PDP was viewed as a means of overcoming a lack of data resources in their state or system. For example, a participant from a state agency noted:

> We're experiencing, like every other state, resource issues, and so to try to build a new data warehouse from the ground up at this point is a little hard, so we're trying to use tools like this to help us along the way…We see it as a cost of doing business at this point. When you can't afford a multimillion-dollar warehouse, it's a deal.

For the small number of states without a centralized or mature data system, the PDP is viewed as a means to overcome data resource constraints–a cost well worth financial resources–to improve capacity. Similarly, a participant from a college system with no centralized state data system noted that

"institutions are just exhausted with their reporting. They just do so much of that, so I think [PDP and third-party reporting capabilities] that's a value add." The value add for this participant and their institutions lies in PDP's data fields and reporting processes, which allow the system and interested third parties to access institutional data without an additional ask to its institutions.

Recommendation: Beyond considerations of cost, recommended is a focus on organizational capacity and finding ways to reduce perceived burdens associated with duplicative reporting in robust data systems. For states with less robust data systems, leveraging the capacity building aspects of the PDP will be an important lever for promoting and advancing use of the PDP in those systems. For both types of systems, regular communications, with specific resources and recommendations for actions and timing, by the PDP could help improve uptake at the state and institutional level (for example, the PDP could provide a state-level guide for promoting institutional PDP use).

CONCLUSION

The PDP holds potential for informing state agency and system policymaking and decision-making. The strong alignment of PDP data fields to state policy priorities and benchmarking potential are among the most valuable aspects of PDP reporting. Further, the relative ease of reporting to the majority of PDP data fields, especially those that were well aligned, is another strength of the tool. Participants had little to say regarding data field reporting; however, they were eager to share their experiences of working with the data, including the ARF and dashboards that PDP reporting provided to them.

For state agencies or systems with less robust data systems, the PDP and, in particular, the ARF provide centralized access to information that they would not otherwise have–creating a significant value to the state entity and, ultimately, their stakeholders. However, for those state agencies and systems with more robust data systems, the PDP holds less value and is often duplicative of work they

can already do in-house. As those make up the bulk of postsecondary oversight entities, the long-term impact of the PDP to inform postsecondary policymaking, decision-making, and outcomes in these states will be minimized. As a result of the varied entity types and their varied perspectives on PDP value, there will likely continue to be uneven reporting and use of the PDP. However, attending to data alignment, state context, and organizational capacity has the potential to improve the PDP utility and uptake.

APPENDIX A

Survey Questions: State-Level Postsecondary Data Partnership Participation

1. Please share your contact information: Name/agency/system/email
2. Are you available to meet to discuss any follow-up question we might have based on your responses to this survey?
3. When did your state or system do the following (please list the year and the month, if you know it, next to each item or list n/a if you have not yet done one of these items):
    - First engaged with NSC about PDP participation?
    - Signed a contract with NSC?
    - Began creating files for initial submission?
    - Sent final files for initial submission?
    - Gained access to PDP dashboards?
    - Gained access to PDP analysis ready files?
4. How many estimated hours did it take to create the initial submission file?
5. How many estimated hours do you anticipate it taking to create subsequent files?
6. Has your state/system's PDP participation been grant funded?
    a. If yes, by whom and for how long have you had funding?
7. Have you or will you submit data on behalf of your state's higher education institutions?
    a. If yes, for how many institutions do you submit?
8. Has your agency provided any resources (e.g., guidance, tools, funding, personnel time, technical assistance) to your state/system institutions for PDP data submission?
    a. If yes, please explain what resources you provided and to whom (for instance 2-year, 4-year, technical institutions).
9. Has your agency provided any resources (e.g., guidance, tools, funding, personnel time, technical assistance) to your state/system institutions for PDP analysis ready files or dashboard data use?
    a. If yes, please explain what resources you provided and to whom (for instance 2-year, 4-year, technical institutions).
10. What is the value in reporting to the PDP? Please select all that apply (choice with open-ended option – multi-select)
    - PDP dashboard access
    - Analysis ready file access
    - Better insight into my state's data
    - Early momentum indicators
    - Easier benchmarking
    - Easier reporting
    - Easier data sharing with third-party organizations
    - Provides tools not currently available through my agency/system
    - Provides access to data and information not previously available at the state/system level
    - Other
11. Which fields are particularly challenging to collect or report to the PDP? Please name the field and explain why. A full list of the PDP fields begins on page 16 of the PDP Submission Guide.
12. What are the challenges in reporting these fields to the PDP? (choice with open-ended option - multi-select)
    - Data not collected/available
    - Different data definitions

- Relying on individual institutions in my state to submit
- Low priority given other demands
- Not enough resources (staff, time, funding)
- Lack of technical assistance or support
- Other

13. Do current PDP dashboards connect to your state policy priorities?
    a. If yes, please explain which fields and priorities and how.
    b. If no, please explain why not.
14. Do you use the PDP dashboards to inform your work? If you have not yet submitted, do you anticipate using the dashboards to inform your work?
    a. If yes, which PDP dashboards do you use in your work?
        - Enrollment
        - Financial aid: unmet need
        - Early Momentum Metrics
            - Credit accumulation rate
            - Credit completion ratio
            - Gateway Course Completion
        - Outcomes over time
            - Outcomes
            - Retention/persistence
            - Retention/persistence term to term
            - Transfer
            - Time to Credential/Credentials Conferred
    b. If you do not use the dashboards in your work or do not anticipate using the dashboards once you have access, please explain why.
15. Do you use or are you planning to use the analysis ready file to inform your work?
    a. If yes, what projects or questions have you used it to answer?
    b. If you do not use the analysis ready file in your work or do not plan to use it once you have access, please explain why?
16. Please share any additional feedback you have related to reporting data to the PDP and using PDP's dashboards and analysis ready file.

APPENDIX B

Table 1. PDP State- and System-Level Participants

Complete List as of August 2021	State ARF Data Shared with SHEEO
Alaska: University of Alaska System	
Colorado: Colorado Community College System	
Connecticut: Connecticut State Colleges and Universities	X
Georgia: University System of Georgia	X
Hawaii: University of Hawaii System	X
Kentucky: Kentucky Council on Postsecondary Education	
Montana: Montana Commissioner of Higher Education	
Nevada: Nevada System of Higher Education	X
New Hampshire: Community College System of New Hampshire	
New York: CUNY System	X
North Carolina: North Carolina Community College System	X
South Carolina: South Carolina Technical College System	
South Dakota: South Dakota Board of Technical Education	
Tennessee: Tennessee Board of Regents	X
Vermont: Vermont State Colleges	

APPENDIX C

Table 2. State- and System-Level Reporting to PDP Data Fields[8]

PDP Data Fields	Total Percentage of Data Field Missingness	Breakdown Percentage of Data Field Missingness: Unknown (UK), Not Applicable (NA), Missing (M), or Blank (B)[2]
Dual And Summer Enrollment	98.40	98.40 UK
Gateway Math Grade – Year 1	95.26	0.02 M; 95.24 B
Gateway English Grade – Year 1	93.89	0.02 M; 93.87 B
Developmental English Completed – Year 1	91.79	91.32 NA; 0.47 B
Developmental Math Completed – Year 1	86.40	85.93 NA; 0.47 B
Gateway English Course Completion – Year 1	86.24	86.24 B
Gateway Math Course Completion – Year 1	86.24	86.24 B
Developmental English Course Attempt – Year 1	85.58	66.95 NA; 18.16 UK; 0.47 B
Gateway Math Course Attempt – Year 1	84.18	83.71 UK, 0.47 B
Gateway English Course Attempt – Year 1	83.82	83.35 UK; 0.47 B
Gateway Math Status	78.01	78.01 UK
Gateway English Status	78.01	78.01 UK
Developmental Math Course Attempt – Year 1	74.61	56.11 NA; 18.03 UK; 0.47 B
Time-To-Credential	73.04	73.04 B
First-Generation Identification	71.73	71.73 B
Program of Study – Term 1	42.73	42.73 B
English Placement	18.50	18.50 UK
Math Placement	18.36	18.36 UK
Pell Status – Year 1	17.79	17.79 UK
GPA Group – Term 1	17.10	17.10 UK
Gender	10.30	10.30 UK
Ethnicity	2.51	2.51 UK
GPA Group – Year 1	1.85	1.85 UK
Credential Type Sought – Year 1	1.84	1.84 UK
Race	1.66	1.66 UK
Enrollment Intensity – Term 1	1.53	1.53 UK
Credits Attempted – Year 1	1.09	1.09 B
Credits Earned – Year 1	1.09	1.09 B
Credits Attempted – Year 2	1.09	1.09 B
Credits Earned – Year 2	1.09	1.09 B
Credits Attempted – Year 3	1.09	1.09 B
Credits Earned – Year 3	1.09	1.09 B
Credits Attempted – Year 4	1.09	1.09 B

[8] Unknown (UK) = data unknown by submitting parties; Not Applicable (NA) = data deemed inapplicable by submitting parties; Missing (M) = submitting parties indicate that a grade is missing from their Gateway Math and English Year 1 data; Blank (B) = no data were submitted to this field.

Credits Earned – Year 4	1.09	1.09 B
Institution ID	0.00	
Cohort	0.00	
Student ID	0.00	
Cohort Term	0.00	
Student Age	0.00	
Enrollment Type	0.00	
Attendance Status – Term 1	0.00	
Retention	0.00	
Persistence	0.00	
Years to Bachelor's at Cohort Institution	0.00	
Years to Associate or Certificate at Cohort Institution	0.00	
Years to Bachelor's at Other Institution	0.00	
Years to Associate or Certificate at Other Institution	0.00	
Years of Last Enrollment at Cohort Institution	0.00	
Years of Last Enrollment at Other Institution	0.00	