2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #52591

Enhancing Faculty Development through Academic Assessment: Moving Beyond Traditional Metrics in STEM Higher Education

Dr. William C. Long, Louisiana Tech University
Dr. Krystal Corbett Cruse, Louisiana Tech University

Dr. Krystal Corbett Cruse is the First-Year Engineering Programs Coordinator and Assistant Professor in the Mechanical Engineering Department at Louisiana Tech University. She is also the Director of the Office for Women in Science and Engineering at Louisiana Tech.

Dr. Mary E Caldorera-Moore, Louisiana Tech University

Dr. Mary Caldorera-Moore is the Associate Dean of Undergraduate Studies, College of Engineering and Science (COES), Associate Professor, Biomedical Engineering, and Daniel D. and Linda D. Reneau Endowed Professorship at Louisiana Tech University.

Enhancing Faculty Development through Academic Assessment: Moving Beyond Traditional Metrics in STEM Higher Education

This research paper is an exploratory study submitted to the Faculty Development Division of the American Society of Engineering Education.

The Grade Point Average (GPA) is a widely used metric in higher education to evaluate student performance, with applications ranging from scholarship eligibility and course prerequisites to program enrollment and graduation requirements. Employers also frequently consider GPA when assessing candidates for internships or job opportunities. Additionally, grade distribution is often used by instructors and administrators to evaluate teaching effectiveness and faculty performance. However, comparing distributions across courses can be challenging, leading to an over-reliance on simplified metrics like the DFW rate, which reflects the percentage of students receiving grades of D, F, or withdrawing (W). While useful, this metric has limitations. For instance, not all courses use D as a passing grade, and the DFW rate doesn't capture the full range of student performance (e.g., whether students predominantly earned As or Cs). While substantial research highlights both the strengths and shortcomings of GPA as a predictor of academic success, one notable limitation is its failure to account for courses from which students withdraw before the drop deadline.

At Louisiana Tech University, the College of Engineering and Science (COES) offers thirteen undergraduate degree programs: seven in engineering, two in engineering technology, and four in math and science. Students are allowed to attend two-thirds of a course during a term before formally withdrawing, which can significantly affect the DFW grade distribution for courses in COES. This study explores an alternative metric that accounts for withdrawn courses to provide a more comprehensive assessment of student performance in particular courses. This alternative metric is a modified form of the GPAW metric as developed by Manteufel et al.[1] and is applied to a course grade distribution to provide a more granular measure of a course's grade distribution to better assess student performance.

This paper additionally presents the findings of a survey conducted among instructors teaching courses within the COES at Louisiana Tech University. The survey collects detailed insights into instructors' perspectives on several key factors: the overall quality of students, the perceived difficulty of the courses they teach, and their assessment of the effectiveness of their instructional methods. With an approximately 57% participation rate, the survey results are analyzed in relation to the alternative metrics for evaluating student performance and course grade distribution. The result showcases the course metric versus student quality, with instructor opinions integrated into a plot. By mapping instructor opinions to these alternative measures, the study highlights how factors like course difficulty and instructional effectiveness correlate with the alternative measures. This approach offers a different perspective for understanding academic outcomes, moving beyond traditional GPA and DFW metrics. The findings provide a foundation for refining course design, instructional practices, and performance assessments, with implications for enhancing both teaching effectiveness and student success in COES.

Background and Context

If you cannot measure it, you cannot manage it. The authors would wager that those in the STEM higher education field would align more closely with this statement than not. As faculty in the STEM higher education field, our development as effective educators must come from a perspective of comparison with our past state or position. That is to say, in order to understand development, we must be judged or measured. In the STEM higher education field, the measuring stick often gravitates to more quantifiable metrics due to our affinity for numbers. However, not all measuring sticks have to be quantifiable. Berk[2] examines 12 common sources of measuring teaching effectiveness. These sources can broadly be grouped into who provides the evaluation:

- Students
- Instructor (self)
- Third-Party (Peers/Administrators/etc.)

Student evaluation of teaching effectiveness comes in many forms, but the most familiar is likely the end-of-course survey. Medina et al.[3] examine the design of these surveys within the context of pharmaceutical education and reviewed over 800 articles to establish a framework focusing on how the survey is designed, administered, and interpreted. They provide recommendations on these three key areas to aid in the creation and implementation of these types of surveys. Other common forms of student evaluation of teaching effectiveness are through student exit interviews. These exit interviews may occur upon completion of a degree, or once a student has decided not to pursue a degree any longer. Kacius et al.[4] discuss how the exit interview is a tool in the continuous quality improvement process. They reviewed over a decade of exit interviews in the public health field to develop a ten-step interview process. Their recommendation includes ensuring anonymity in feedback, linking the interview to graduation requirements, and, for large programs, using sampling. Other forms of measuring teaching effectiveness include student evaluation assessing student performance using metrics such as grades in a particular course or follow-on courses[1]. However, not much research has been devoted to this space [11].

Instructor evaluation of teaching effectiveness presents itself in a number of ways, mostly ranging from self-reported items such as self-evaluation, teaching portfolios, or even video self-confrontation[2]. Villanueva et al.[5] examine the use of various teaching evaluation practices across the country to determine which practices engineering faculty report as useful. Of the practices reported, informal self-reflection proved to be a meaningful way of evaluation. Simonson et al.[7] describe the teaching portfolio as a compilation of evidence that is more appropriate than any single source of information. With this approach, the teaching portfolio requires the instructor to gather evidence that demonstrates achievement. Simonson et al. recommend including student work, course material, and rationales for instructional choices[7]. Herman et al.[6] discuss the use of video recordings of teaching as a way for improving teaching practices. Their recommendations include a private review of the video with a written reflection summarizing the viewing, including strengths and specific areas for improvement.

Third-party evaluation of teaching effectiveness is the third broad group of evaluation sources and may include peers, administrators, or external entities. Third-party evaluation is commonly used in various contexts, including annual evaluations and promotion dossiers. Godbout-Kinney et al.[8] discuss the shortcomings of student evaluations of teaching and promote the use of third-party evaluations to reduce the known bias student evaluations of teaching have against women and minority faculty members. Quenette et al.[9] also promote the use of third-party evaluations, stating that peer review facilitates learning from colleagues and other experts. The thought is that a third party is better equipped to evaluate an instructor's knowledge of the course material, the rigor of the course, and the quality of prepared materials. Much of the discussion of third-party evaluation of teaching effectiveness in the literature revolves around the distinction between formative and summative evaluation[2], [5]-[9].

Recognizing the scarcity of research on the relationship between student performance and teaching effectiveness, we propose a method to evaluate teaching effectiveness that uses an easily calculated, quantifiable metric similar to a GPA calculation. This approach, based on course outcomes and student performance, is open to some interpretation by the evaluator or the individual reviewing the data to facilitate formative evaluation. The method is further paired with self-reflection faculty survey data to potentially incorporate all three evaluation sources (students, instructor, and third-parties), resulting in a well-balanced yet meaningful evaluation metric.

Method and Rationale

The process begins by utilizing Manteufel et al.[1] metric of *GPAW*. The mean *GPAW* is defined by Manteufel et al.[1] as

$$GPAW = \frac{4n_A+3n_B+2n_C+1n_D+0n_F+0n_W}{n_A+n_B+n_C+n_D+n_F+n_W}$$

where n_X is the number of As, Bs, Cs, Ds, Fs, or Ws (withdrawals) a student has or the number of a particular letter grade from a course's grade distribution. Manteufel et al.[1] postulate that Ws (withdrawals) should be included in a course's grade distribution, as this effectively counts as a failed attempt in a particular course. While there may be numerous reasons a student decides to withdraw from a course, students who habitually withdraw will not progress academically, indicating a lack of achievement of course knowledge. Manteufel et al.[1] use this metric to test the null hypothesis that there is no difference in grade distributions across statics courses taught by multiple instructors. Manteufel et al.[10] further use this metric in mentoring adjunct faculty by comparing the *GPAW* of similar courses or historic values for a particular course. The alternative, and often more prevalent, approach is the use of a DFW percentage rate. DFW rates have limitations and often oversimplify a course's outcomes. For example, consider two grade rosters: one consisting entirely of As, and the other entirely of Cs. Both of these grade rosters would have a DFW percentage of 0, but the *GPAW* values would be 4.0 and 2.0, respectively. This difference paints a very different picture between the two hypothetical courses. With this in mind, the authors developed an approach that utilizes the *GPAW* metric in two different contexts. The first context is based on a course's grade roster, and is designated as the *course GPAW*. The second context is based on a course's student roster. For each student on a course's roster, a *GPAW* value is calculated based on that student's academic history (i.e., their transcript). Since this is meant to be a formative method or approach, for this particular study, the current term in which a student is enrolled in a particular course is included in this academic history calculation. That is to say, the values used in the student's *GPAW* calculation include all previous terms and the term in which the particular course was taken. This approach can only be utilized after the completion of the term and thus cannot be used as a predictive model. Once the *GPAW* values for all students in a particular course are determined, the mean of the students' *GPAW* values can be used to compute the *cohort GPAW* for a particular course. With these two metrics defined, a plot similar to that shown in Figure 1 can be created.

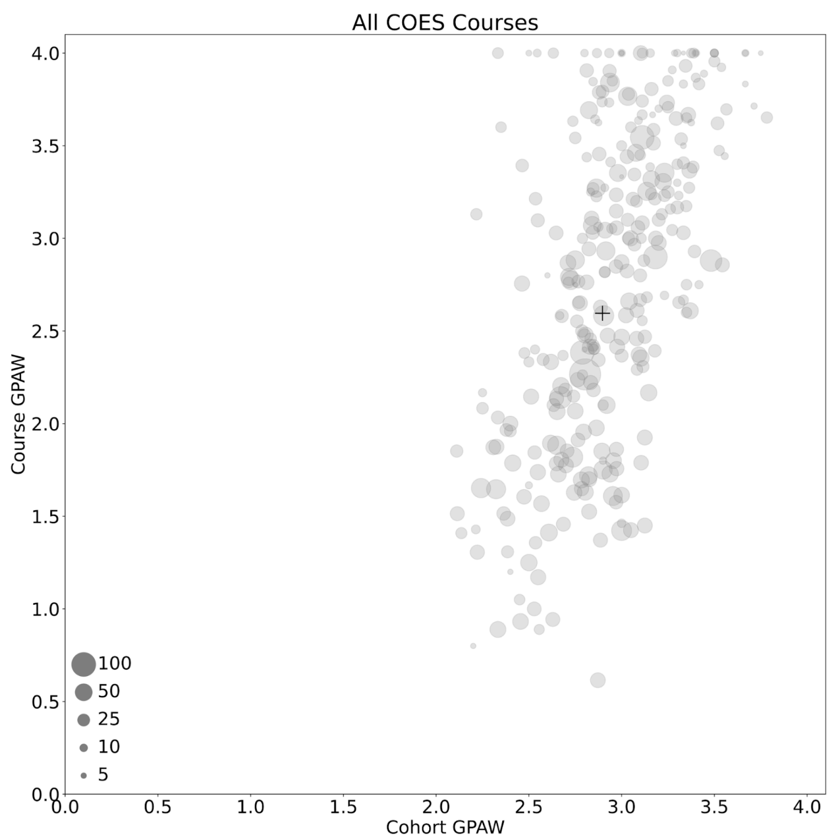

Figure 1. A *GPAW* plot for all COES courses. Enlarged view found in the appendix.

Depicted in Figure 1 are all courses from the COES at Louisiana University for the Fall term of 2024. The relative size of the marker indicates the course roster size (a legend for the marker size is shown in the bottom left corner of the plot), and the plus sign marker indicates the mean value of all the courses shown. When viewed from this perspective, the plot may be interrupted in various ways. Broadly speaking, the *cohort GPAW* axis represents the quality of students in a particular course, and the *course GPAW* axis represents the outcomes of a course. If this overly simplified view is taken, one would conclude that the markers should lie along a 1-1 line. The results shown in Figure 1 indicate otherwise, but a slight trend may be discerned. Further review of the plot yields three distinct regions of note. They are depicted in Figure 2, shown below.

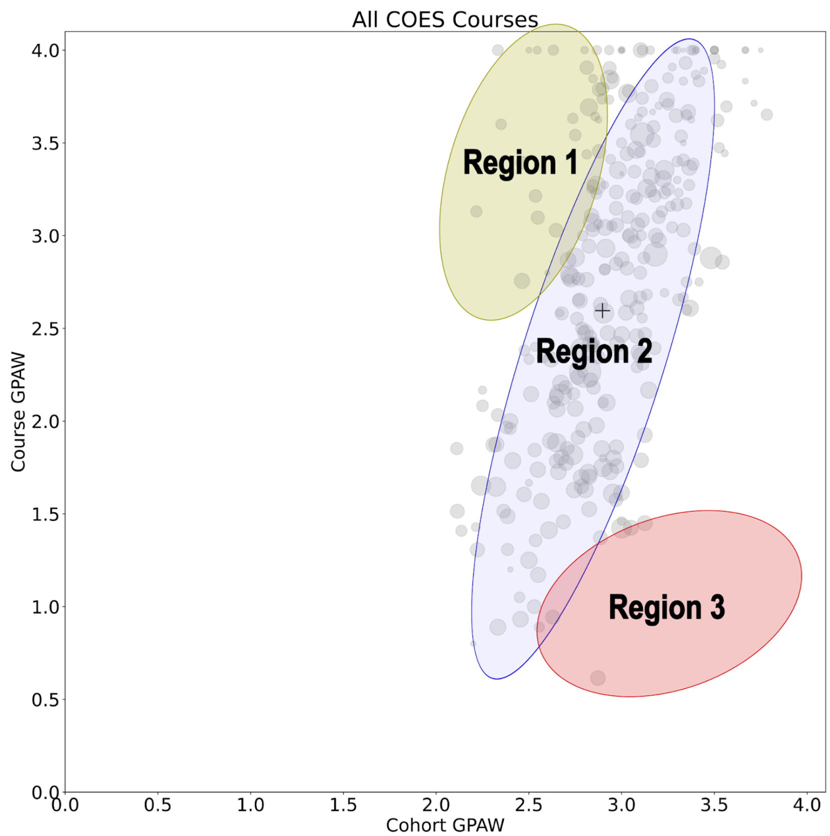

Figure 2. A *GPAW* plot with key regions highlighted.

Region 1 shows courses with positive or high outcomes, but student quality is lower than you might expect given the course outcomes. While there may be a number of reasons for this, the two that come most quickly to mind are that the course subject matter is less difficult than in other courses or that the instructor is inflating the course grades. Region 2 contains the largest group of markers and indicates the general trend observed in the plot. This general trend may be interpreted as normalcy, and deviations indicate a need for closer inspection, as in the case of Region 1 and 3. Region 3 shows courses with low outcomes, but the student quality is high. As in Region 1, there are many possible causes for such a trend, but two are readily reached. Either the course subject material is overly difficult, or the instruction provided is of poor quality. Given that this approach is intended to be formative, conversations with instructors in Regions 1 and 3 would be warranted to better understand what is being observed here.

This study further pairs this metric with a self-reflection survey given to faculty who taught the courses shown in Figure 1. The survey was presented to faculty via printed paper copies that included a grade distribution plot for each course the faculty member taught in the Fall term of 2024. Under each grade distribution, the following question was asked.

> To what extent do you feel your course's grade distribution, shown above, reflects the following aspects?
>
> - The quality of students
> - The difficulty of the course
> - Your effectiveness as an instructor

Each of the bulleted items could be responded to using a five-point Likert scale with the following options.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Results and Discussion

The results presented here represent all undergraduate courses from the COES for Louisiana Tech University for the Fall term of 2024. It includes over 300 courses taught that term, with class sizes ranging from a few to over 150 students with the average class size being approximately 31. The COES has approximately 3000 undergraduate students and includes seven different engineering programs, two engineering technology programs, and four science programs. These courses were taught by approximately 130 different instructors, with various positions including graduate assistant, adjunct professor, visiting lecturer, lecturer, tenure-track professor, and tenured professor. Some of the courses included here represent service courses for the other colleges at Louisiana Tech University while others are taken solely by COES students. The following sections illustrate how the various perspectives of the *GPAW* plot may be used for formative faculty development.

Instructor View

Figure 3 shows a comparison of three instructors and their respective courses, indicated by blue diamond markers.

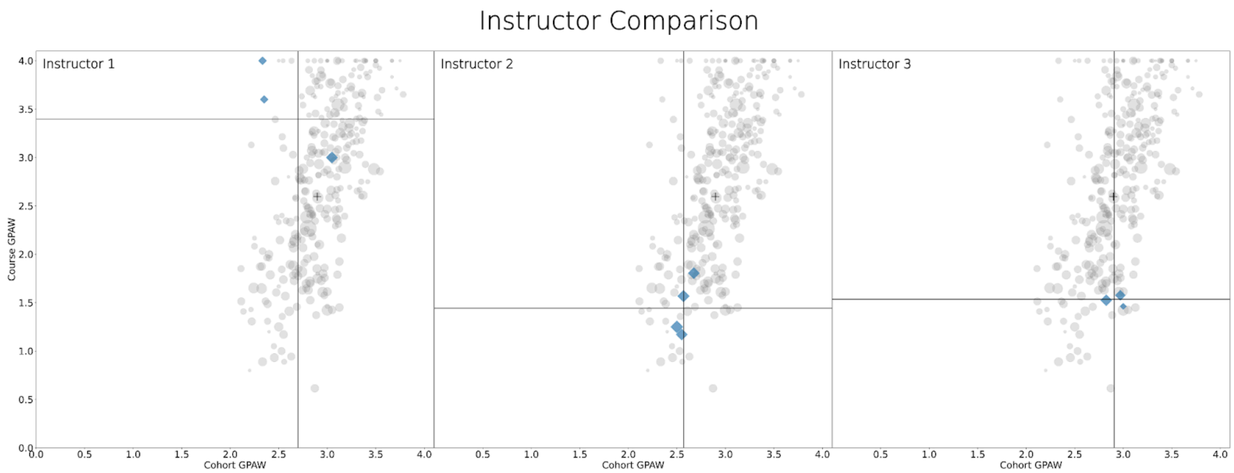

Figure 3. Instructor Comparison of *GPAW* plots. Enlarged view found in the appendix.

The means for the instructor's *cohort GPAW and course GPAW* are represented by the vertical and horizontal lines spanning the plot's height and width, respectively. It may be seen that the three instructors fall largely into one of the three regions previously discussed. Instructor 1's courses fall mostly in Region 1, which could possibly indicate that the course subject material is less difficult or rigorous than other courses or that the instructor is more lenient when it comes to grading. Instructor 2's courses fall largely in Region 2, and illustrate one of the most attractive traits of implementing this approach. The DFW rate of all of Instructor 2's courses is approximately 54%. Without the context of this *GPAW* plot, one might conclude that the instructor's effectiveness is minimal at best; but, given the context of the *GPAW* plot, it may be seen that students who enroll in Instructor 2's courses have a history of lower performance. There may be any number of reasons as to why Instructor 2's students choose to enroll in the instructor's courses. For example, it may be that other sections of the courses Instructor 2 teaches are at more convenient or desirable times, and those who aren't able to schedule for their preferred time must take the available course times (i.e., having to take an 8 AM instead of a 10 AM course). Another possibility is that the instructor has a certain reputation among students, and those who can choose other instructors do so. Anecdotally, the authors have witnessed this play out on several occasions. Instructor 3's courses fall on the border of Region 3 and could possibly indicate that the subject material is difficult or that the instructor is not effective at communicating it. Further discussion with the instructor is warranted. Anecdotally, the authors do note that Instructor 3 has a reputation for being overly difficult, and some students try to avoid taking the instructor's courses.

Subject View

There are other ways of viewing the *GPAW* plots as well. Figure 4 compares different subject areas.

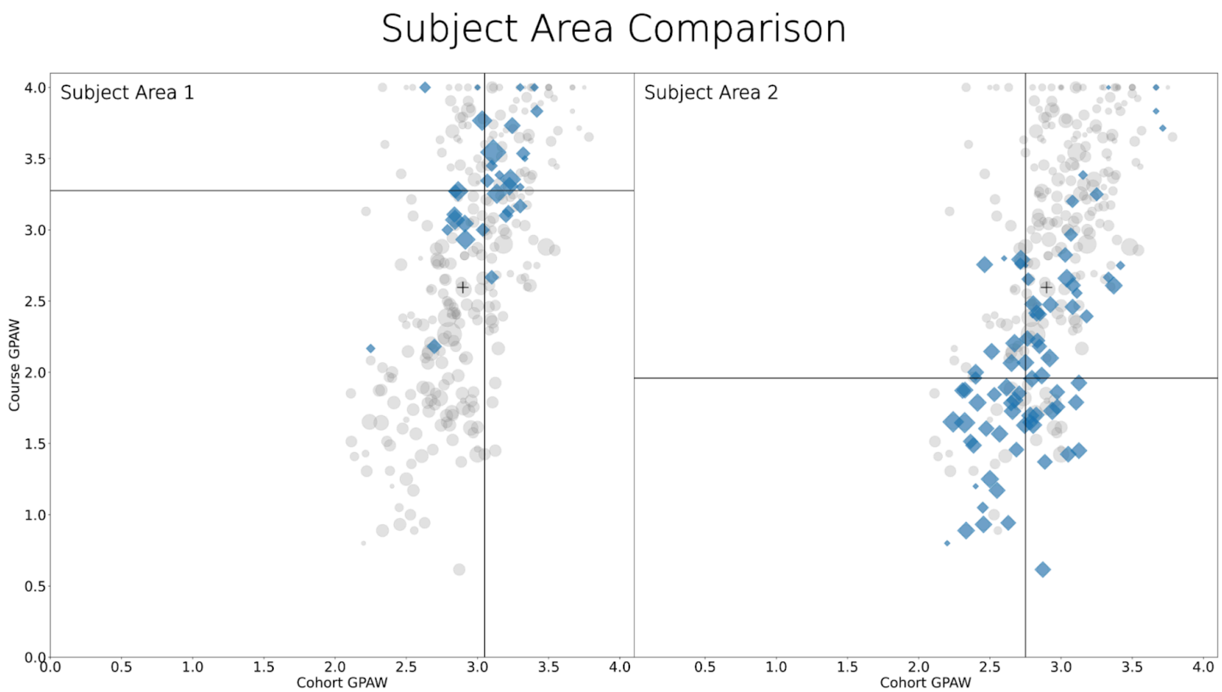

Figure 4. Subject Comparison of *GPAW* plots. Enlarged view found in the appendix.

Figure 4 depicts the *GPAW* plots of two different subject areas in the COES of Louisiana Tech University. Both subject areas represent a science and are largely service courses for the entire university. As in the case of the instructor's view plots, the horizontal and vertical lines spanning the width and height of the plots represent the mean *course GPAW* and *cohort GPAW*, respectively. As can be seen, the grouping centers on two distinct areas and may indicate certain aspects or attributes of the subject area.

Course Level View

More granular views can be created as well. Figure 5 compares the different subject levels (i.e., 100-, 200-, 300-, or 400-level courses) among the COES courses. The mean *course GPAW* and *cohort GPAW* depicted as the horizontal and vertical lines spanning the plot's width and height, show a positive trend, indicating that as subject levels increase, the course outcomes also increase. This trend may be counterintuitive, as higher-level courses are generally thought to be more difficult and to require more prerequisite knowledge for comprehension.

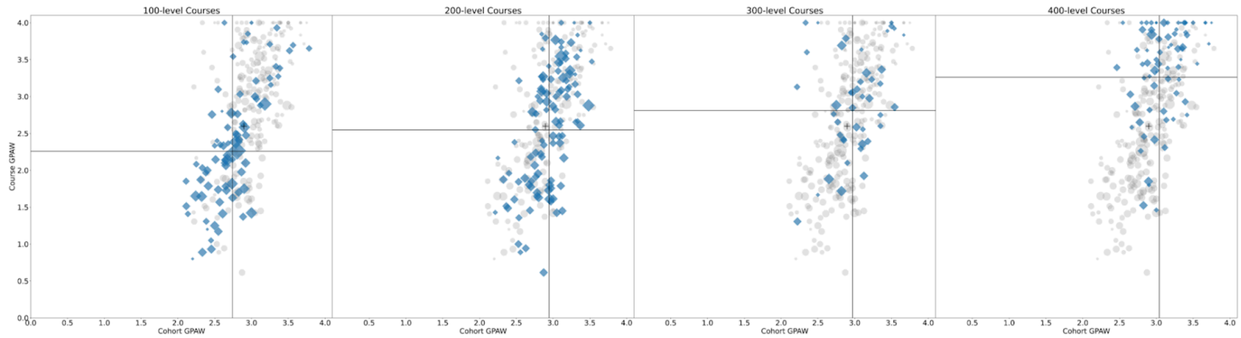

Figure 5. Subject Level Comparison of *GPAW* plots. Enlarged view found in the appendix.

Perhaps therein lies the reason for the positive trend. Since higher-level courses typically require prerequisite knowledge or competency to enroll, students who do not have the necessary abilities or skills to succeed in the course are naturally excluded from registering by way of deficient prerequisite coursework.

Faculty Survey View

The faculty survey portion of this study included the collection of faculty opinions in three key areas: student quality, course difficulty, and instructor effectiveness. Of the approximately 130 instructors who taught undergraduate courses in the Fall term of 2024 for the COES at Louisiana Tech University, 74 responded, yielding a participation rate of about 57%. The faculty responses in these three key areas were further paired with the *GPAW* plot for the courses of the faculty that participated in the survey.

Figure 6 shows the *GPAW* plots with the faculty's opinions about student quality. The majority of faculty marked either Agree or Strongly Agree to the question of “To what extent do you feel your course’s grade distribution reflects the quality of students?” Being that the instructors are, in a large part, the ones determining the grade distribution in their courses, the results of the responses should come as no surprise.

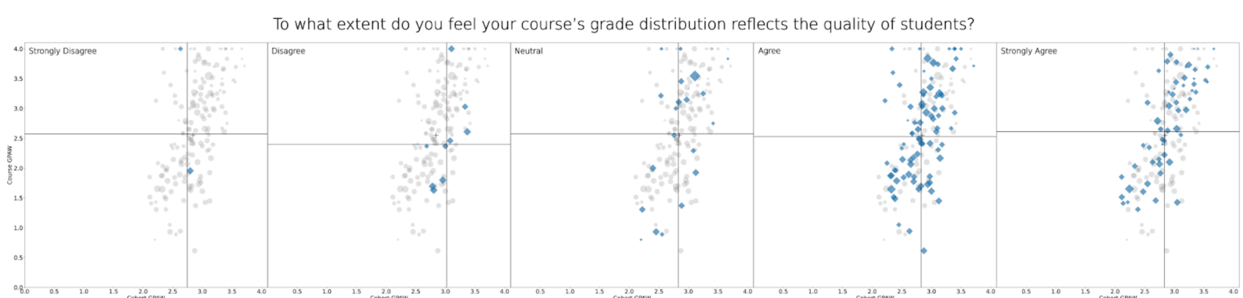

Figure 6. Faculty’s Opinion of Student Quality Comparison of *GPAW* plots. Enlarged view found in the appendix.

Figure 7 shows the *GPAW* plots with the faculty's opinions on course difficulty. Similarly to the previous survey question, the majority of faculty marked Agree to the question "To what extent do you feel your course's grade distribution reflects the course difficulty?" Interestingly, there appears to be a grouping within the courses where faculty marked Disagree. These courses fall lower down on the natural trend of all the courses. Based on the previous analysis of the data, the group's location suggests that these are mostly lower subject-level courses or perhaps even fall within a particular subject area similar to Subject Area 2.

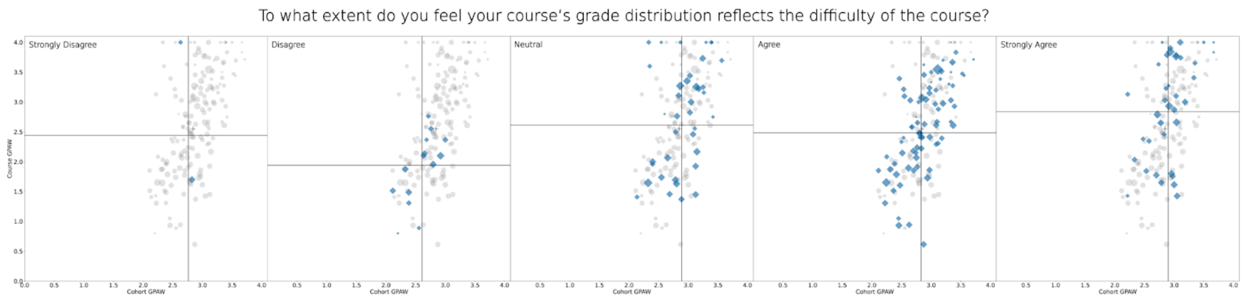

Figure 7. Faculty's Opinion of Course Difficulty Comparison of *GPAW* plots. Enlarged view found in the appendix.

Figure 8 shows the *GPAW* plots for the last question asked on the faculty survey dealing with the faculty's opinions of instructor effectiveness. The question read "To what extent do you feel your course's grade distribution reflects your effectiveness as an instructor?" If the Strongly Disagree group is not considered, there appears to be a slight positive trend between the instructor's opinions about their effectiveness and the mean *course GPAW* values. As the previous questions also indicated, the majority of responses fell into the Agree group.

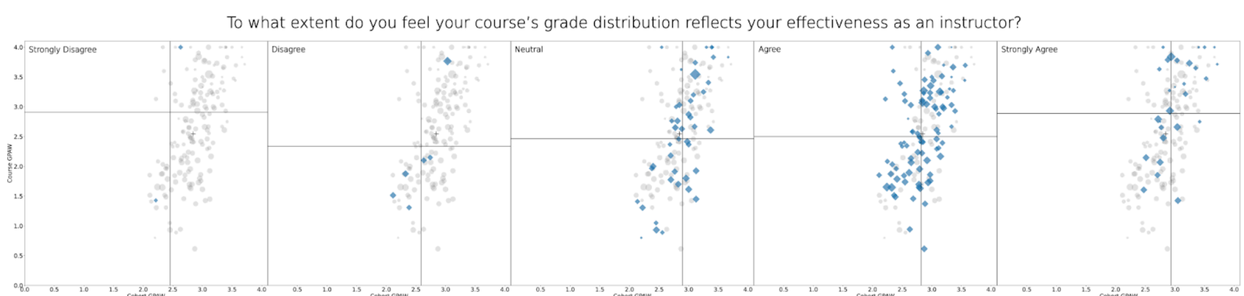

Figure 8. Faculty's Opinion of Instructor Effectiveness Comparison of *GPAW* plots. Enlarged view found in the appendix.

Application

The proposed method is not meant to be a definitive measure of teaching effectiveness, but rather used as a contextualized performance indicator intended to support reflective practice, identify exploratory patterns, and inform targeted faculty development interventions. In the hands of a faculty developer, the method provides another perspective that may be utilized to have conversation starters with instructors. This may take the form of end-of-term course reflections, annual faculty evaluations, etc. For example, the instructor's plots, shown in Figure 3, may be provided to the instructor by a faculty developer as a springboard for learning more about the instructor's methods and practices. The method may also be used as another metric in the development of strategic initiatives of academic units, colleges, etc. For example, the subject perspective plots, seen in Figure 4, may be used in the exploration of understanding the potential lower performance of a particular program or department. The subject-level plots, shown in Figure 5, could provide an administrator with a quick snapshot of the performance of different subject levels. While course subject levels are different than student classifications (i.e., first-year, second-year, etc.), it is reasonable to say that a first-year student will likely be taking more 100-level courses than 400-level courses and vice versa. All of these example applications provide an alternative perspective that may illuminate some underlying factor that might otherwise go unnoticed.

Conclusion

The approach outlined here aims to provide faculty with a different perspective on interpreting course grade outcomes. Utilizing the *GPAW* metric allows for a nuanced view of instructor effectiveness, subject or course difficulty, and the quality of students. The overly simplistic DFW percentage has limitations that the *GPAW* metric, in part, overcomes by offering a more graduated view of the course grades (i.e., a GPA-like value vs. a simple percentage), thus providing more information to those interpreting the data while still being relatively simple to calculate. The *GPAW* metric also offers areas for further research. Several trends were identified throughout this work that deserve investigation to determine whether statistical significance exists. A longitudinal study could be conducted to track trends over multiple terms. Comparing the instructor position (e.g., lecturer, tenure-track, or tenured) or instructor teaching load in the context of *GPAW* plots could also be an interesting perspective. A principal component analysis could be conducted to identify groupings worth investigating. In all, the approach used with this *GPAW* metric should prove useful to any faculty member seeking a better understanding of course outcomes.

Appendix

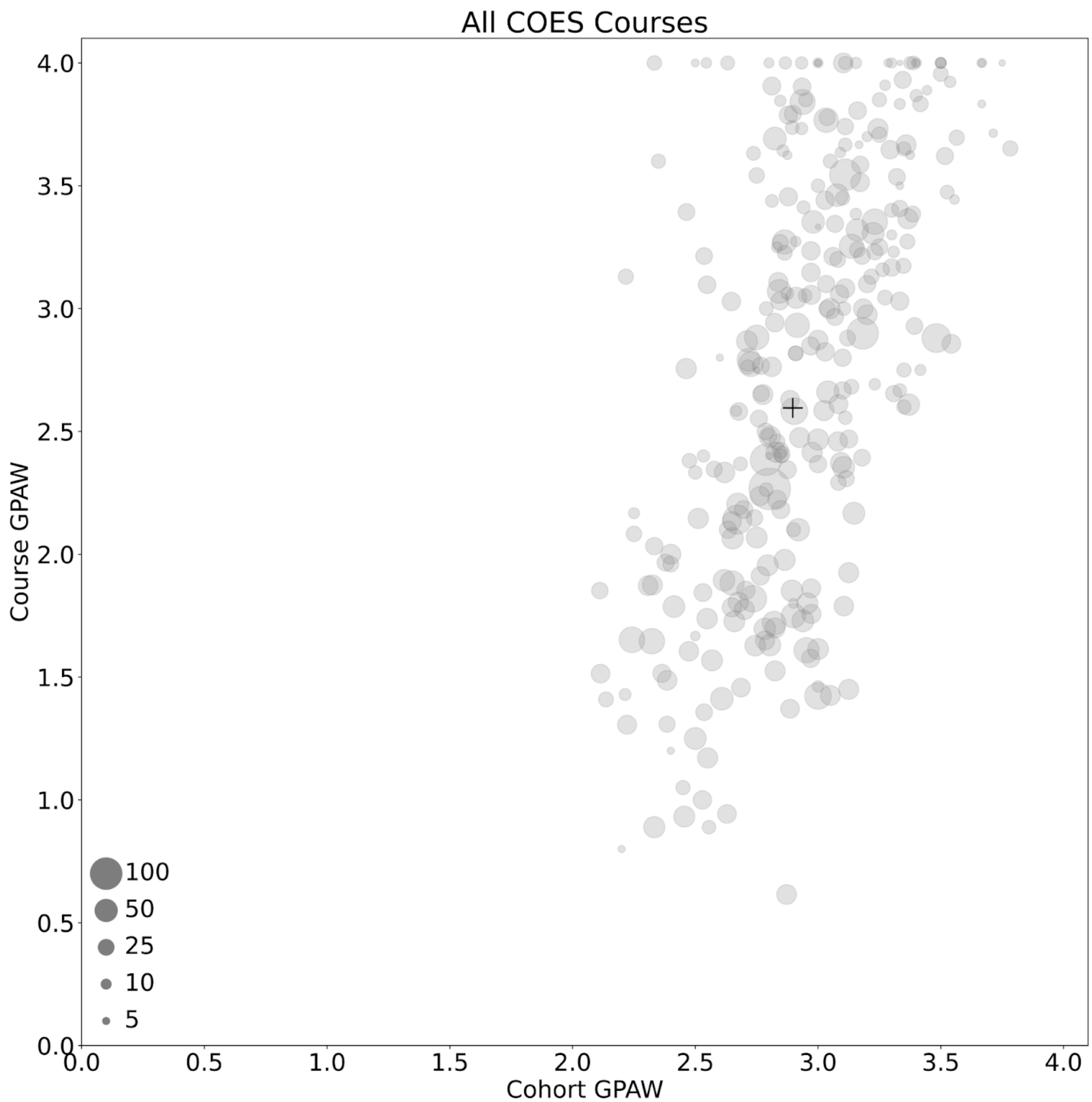

Figure 1. A *GPAW* plot for all COES courses.

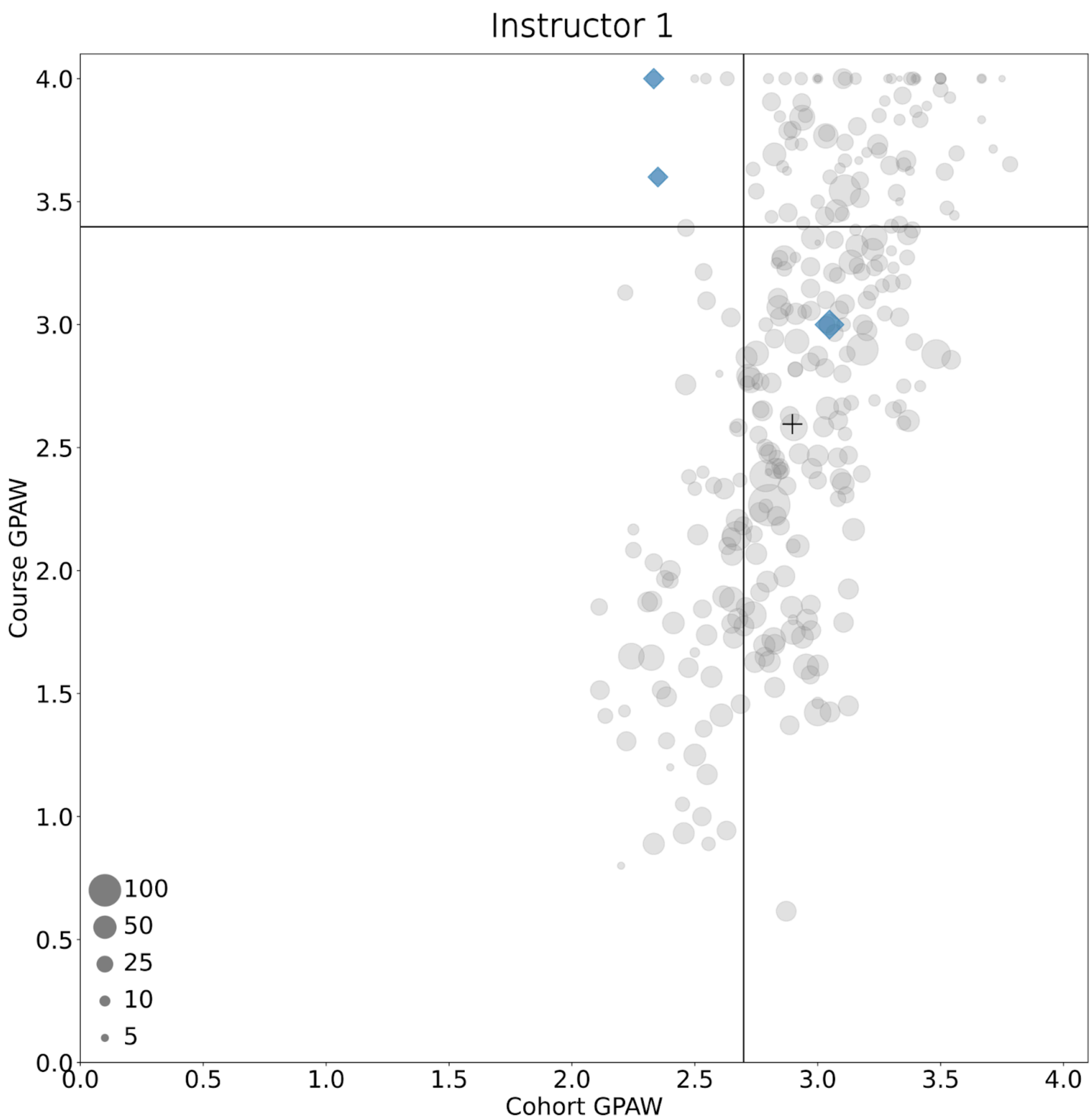

Figure 3.a. Instructor 1 *GPAW* plot.

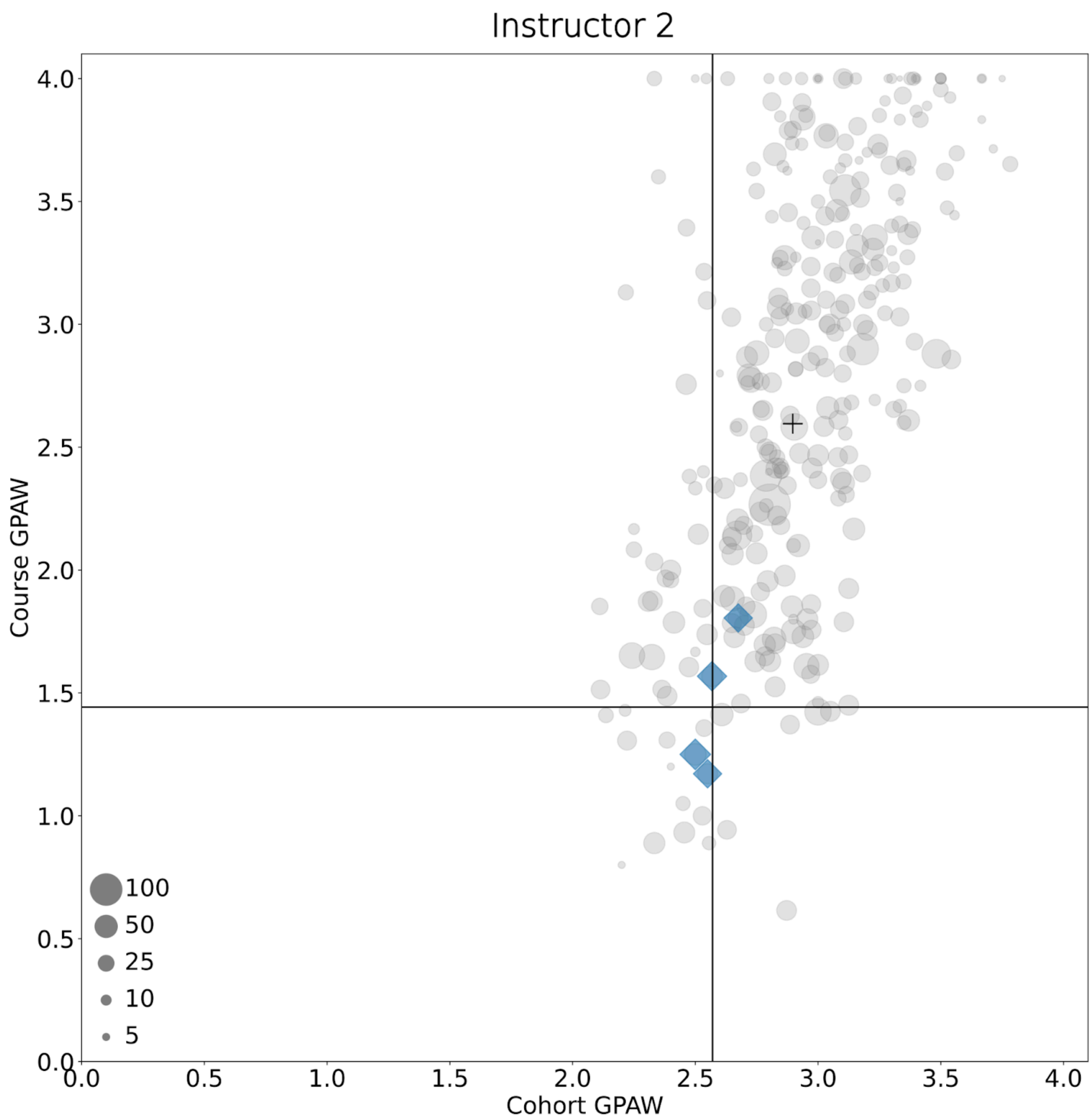

Figure 3.b. Instructor 2 *GPAW* plot.

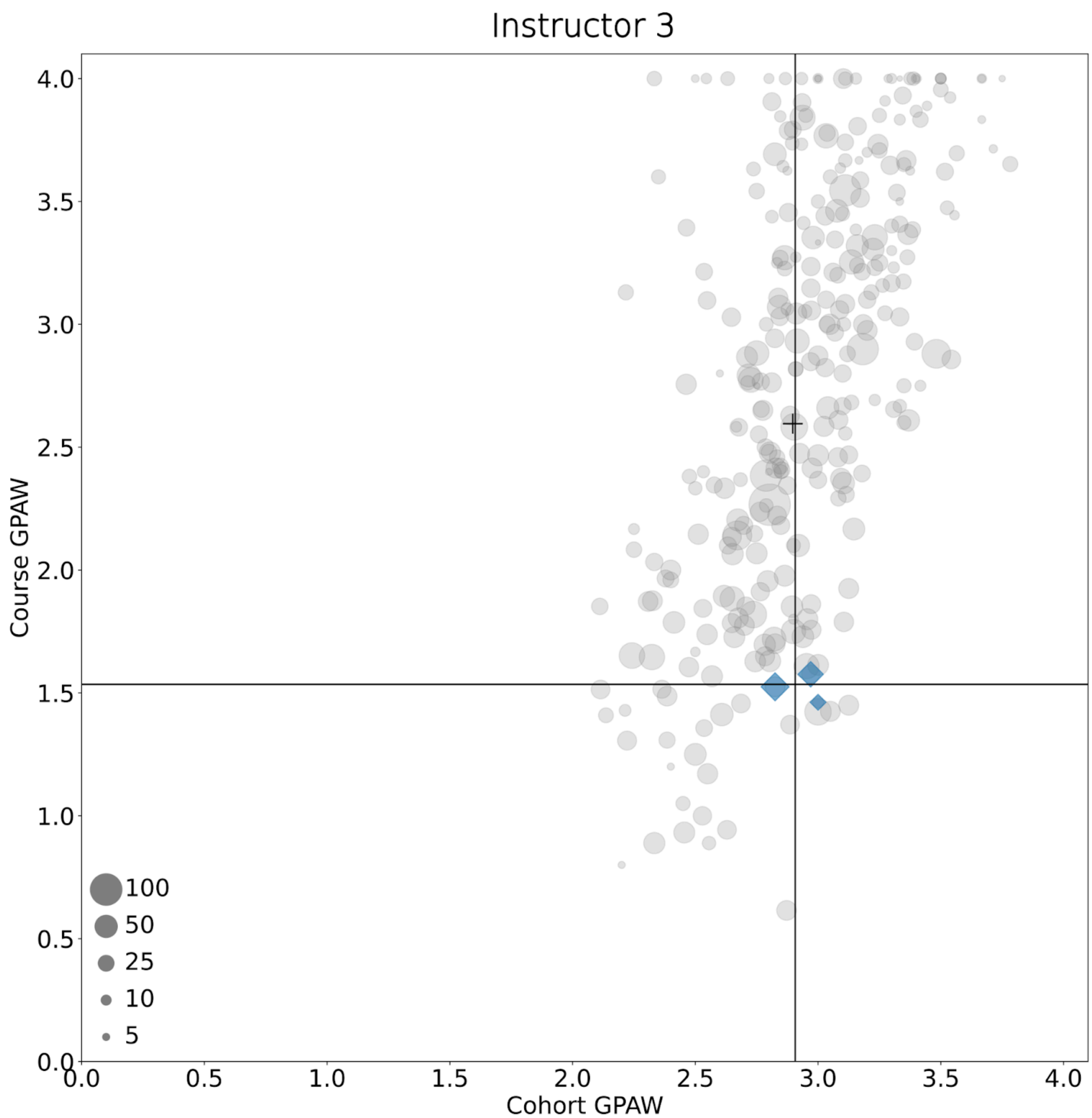

Figure 3.c. Instructor 3 *GPAW* plot.

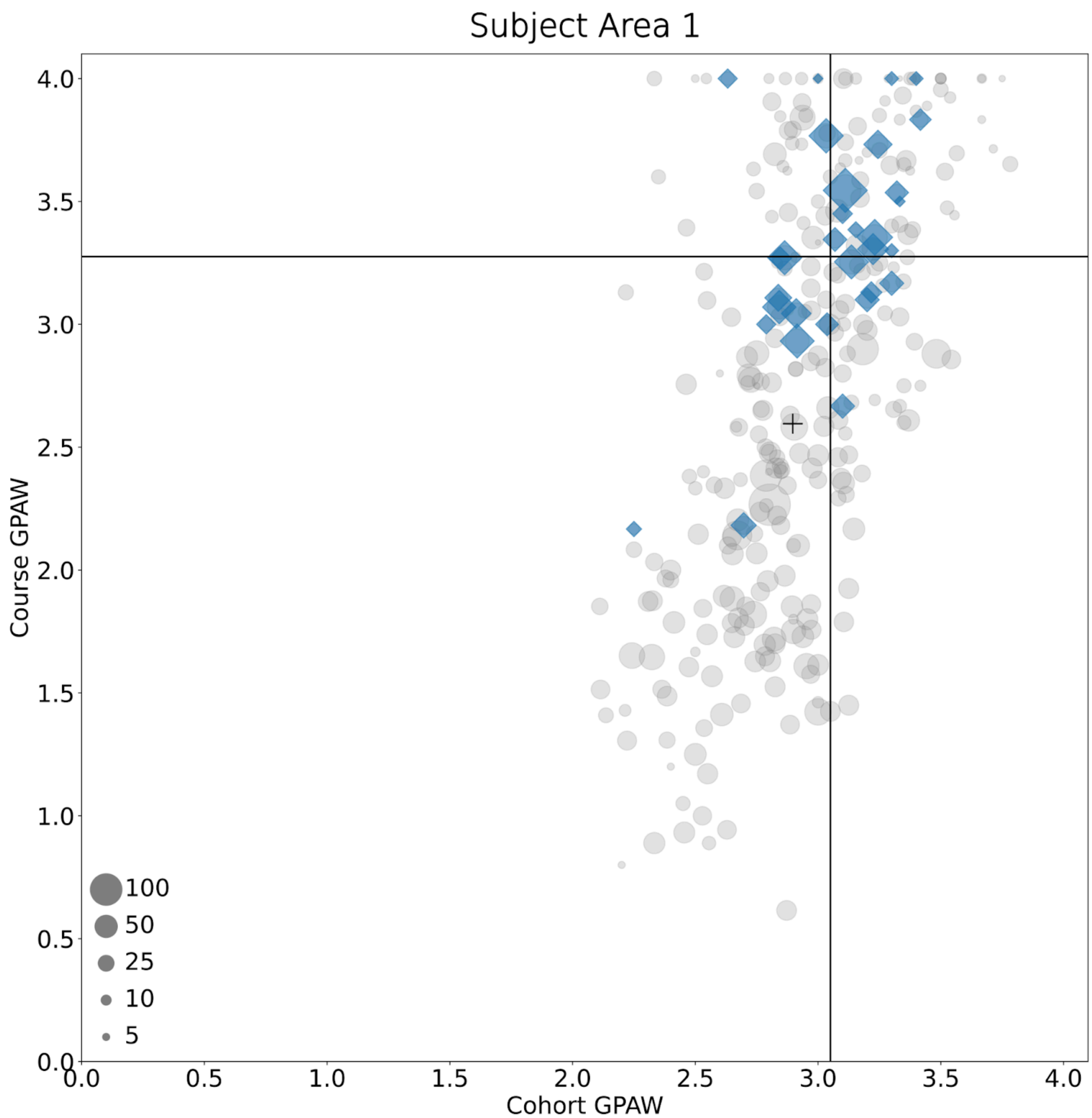

Figure 4.a. Subject Area 1 *GPAW* plot.

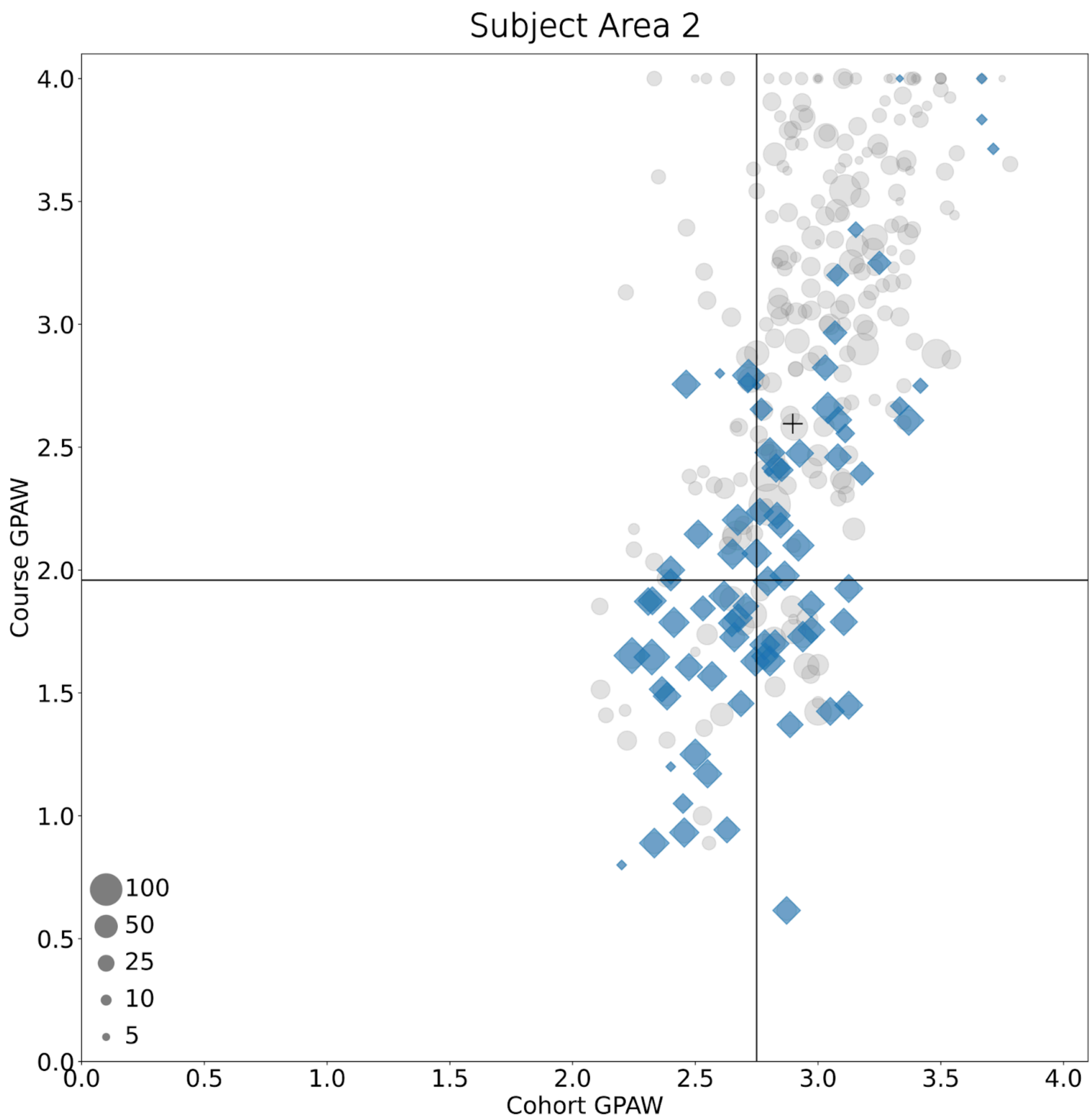

Figure 4.b. Subject Area 2 *GPAW* plot.

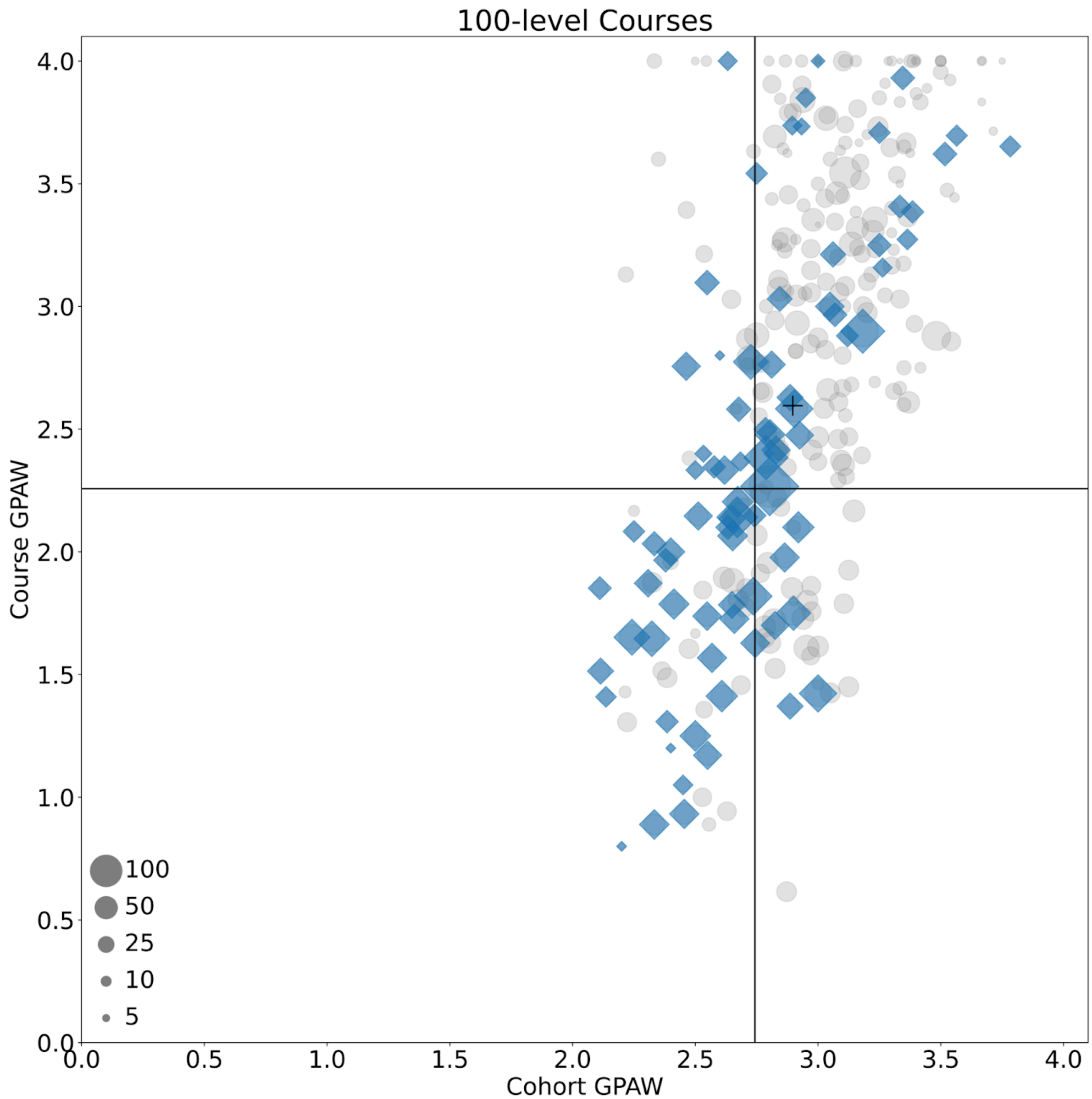

Figure 5.a. 100-Level Courses *GPAW* plot.

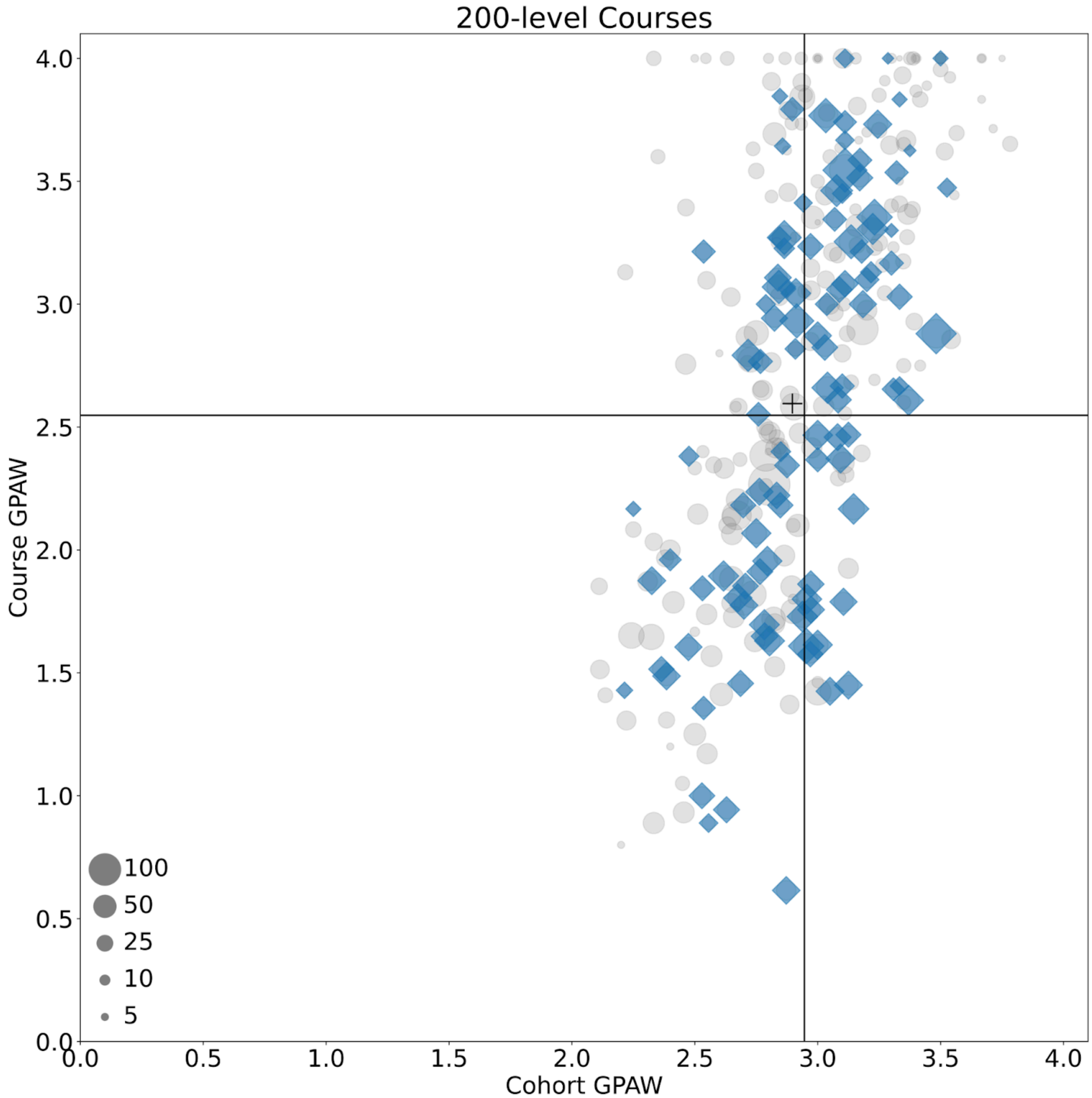

Figure 5.b. 200-Level Courses *GPAW* plot.

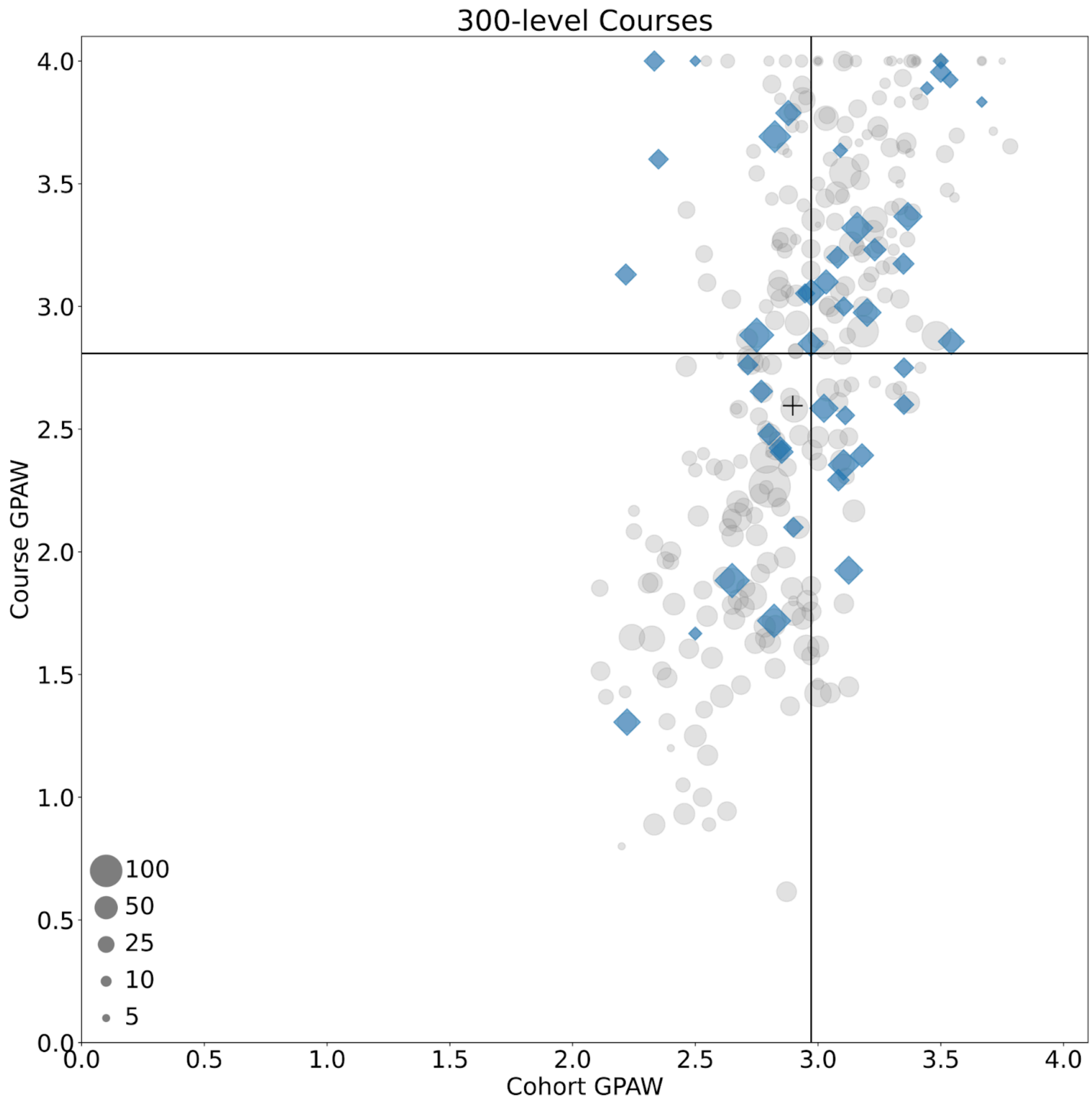

Figure 5.c. 300-Level Courses *GPAW* plot.

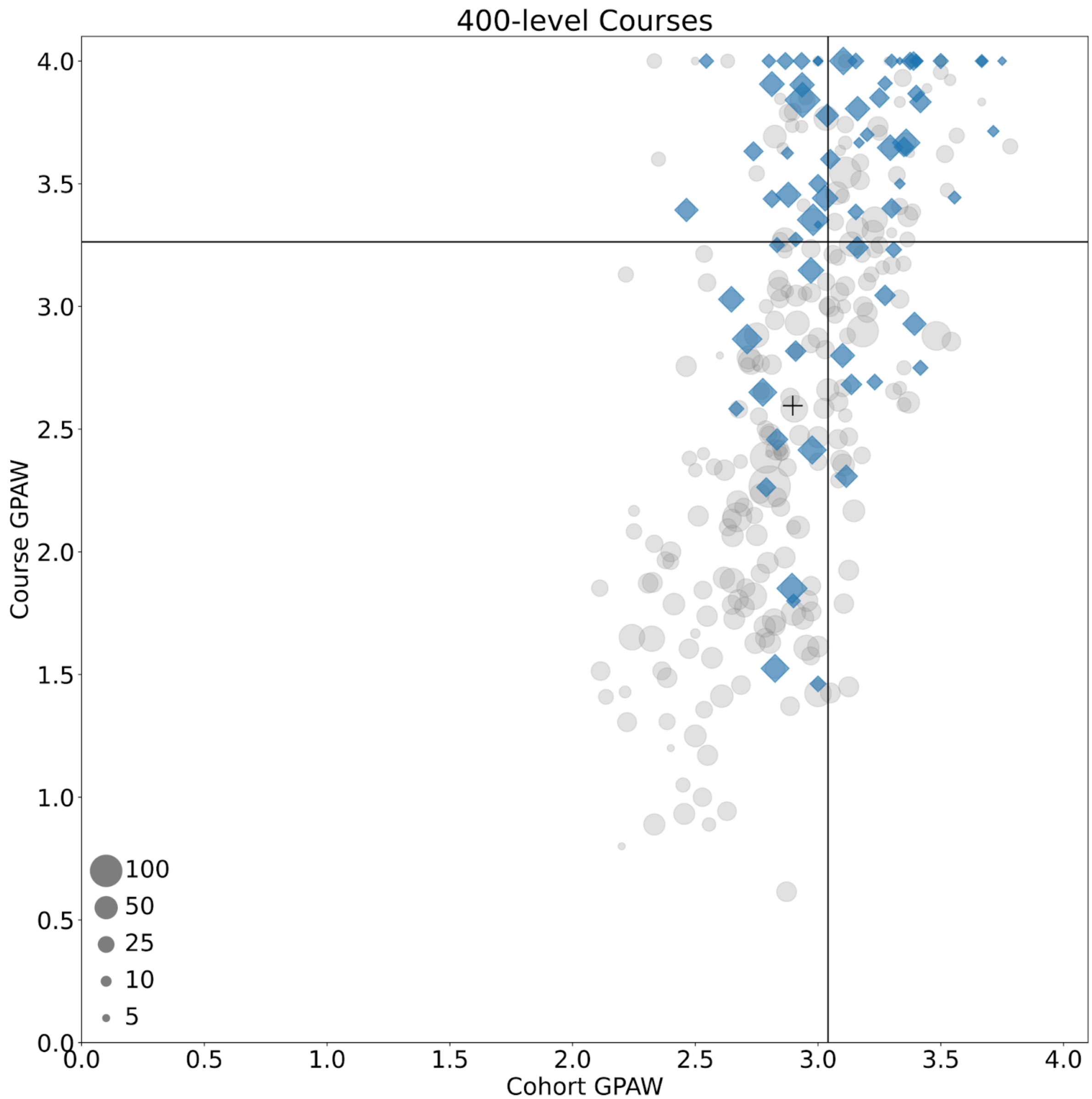

Figure 5.d. 400-Level Courses *GPAW* plot.

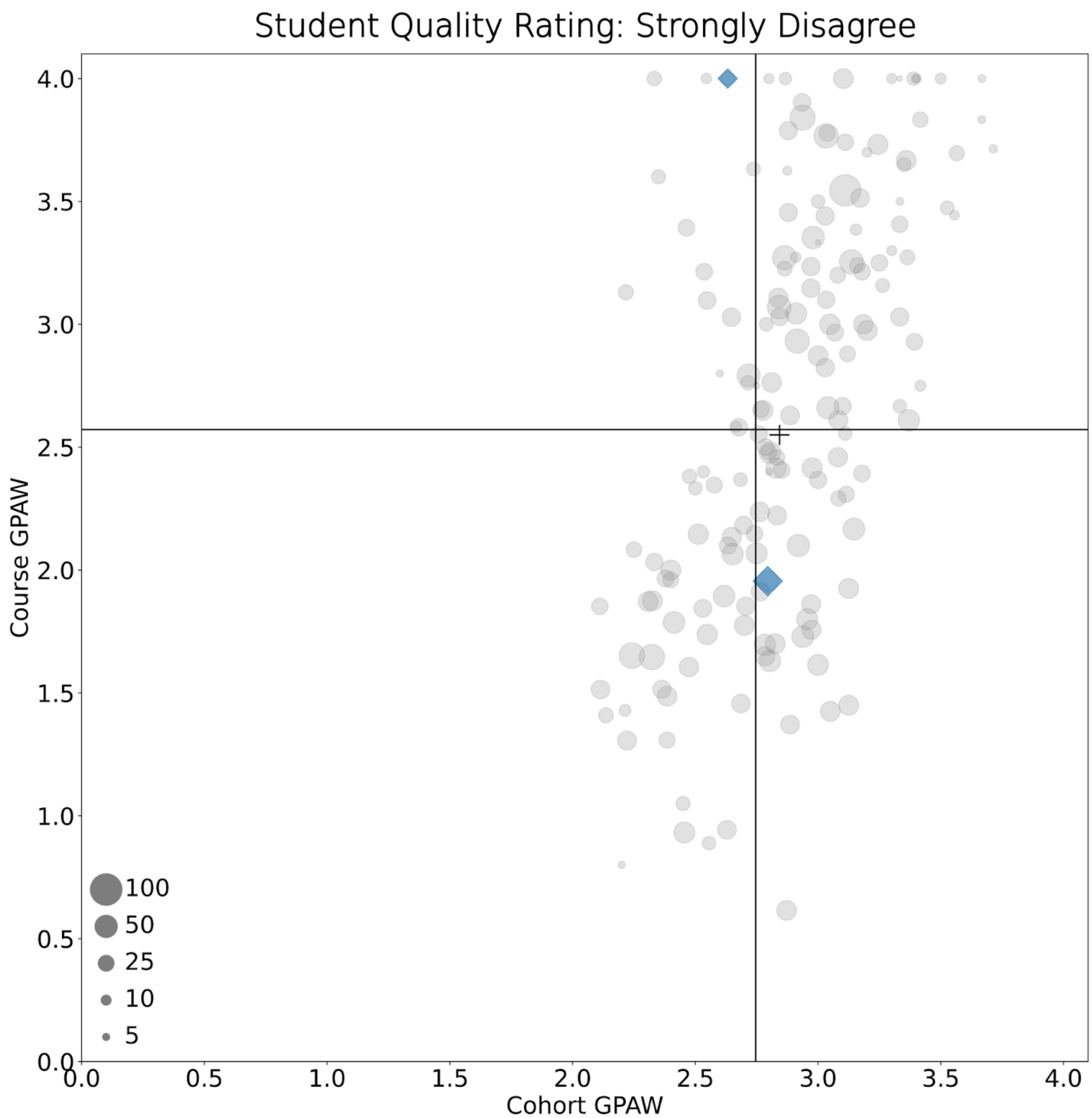

Figure 6.a. Faculty's Opinion of Student Quality - Strongly Disagree Rating.

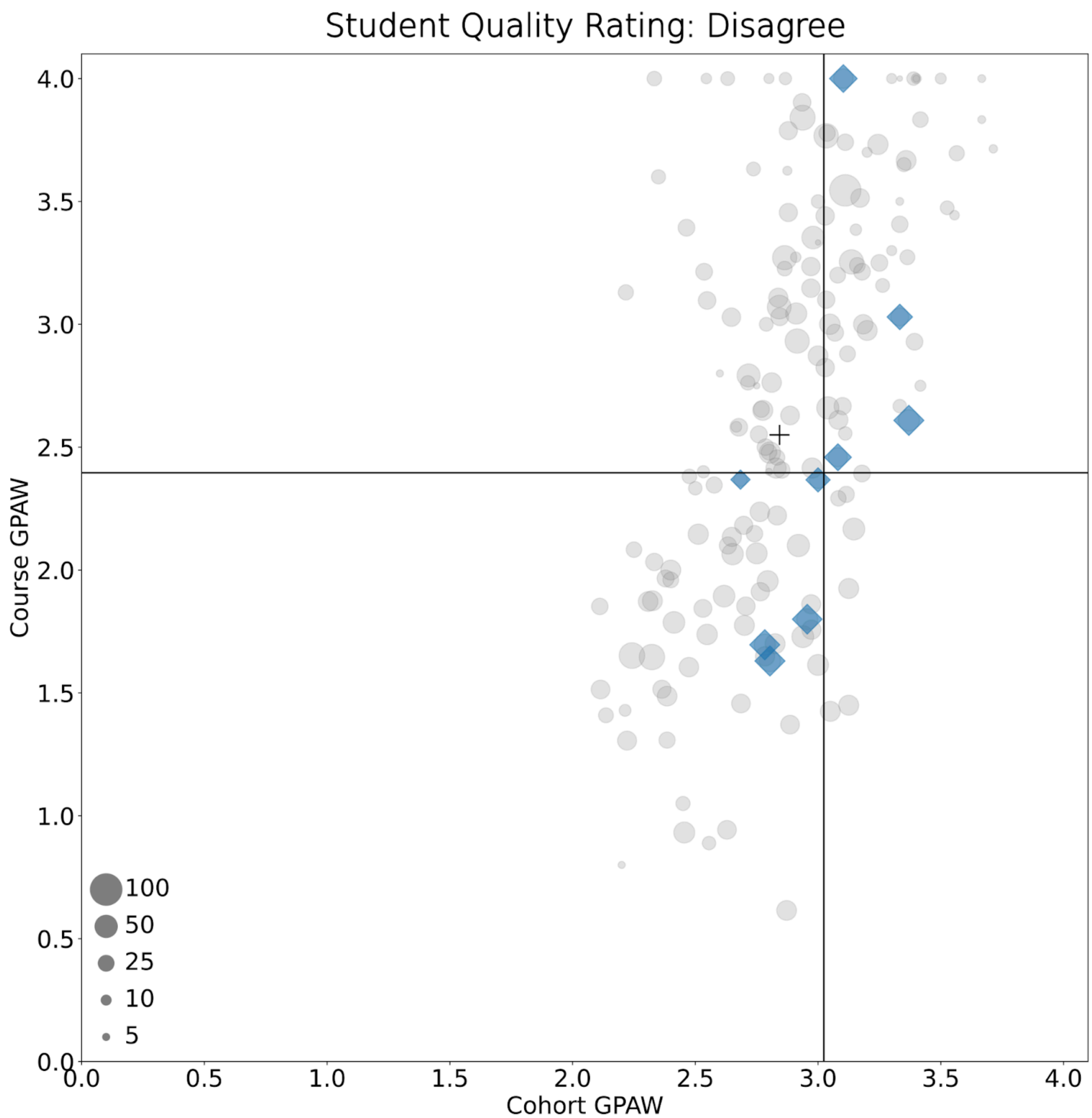

Figure 6.b. Faculty's Opinion of Student Quality - Disagree Rating.

Figure 6.c. Faculty's Opinion of Student Quality - Neutral Rating.

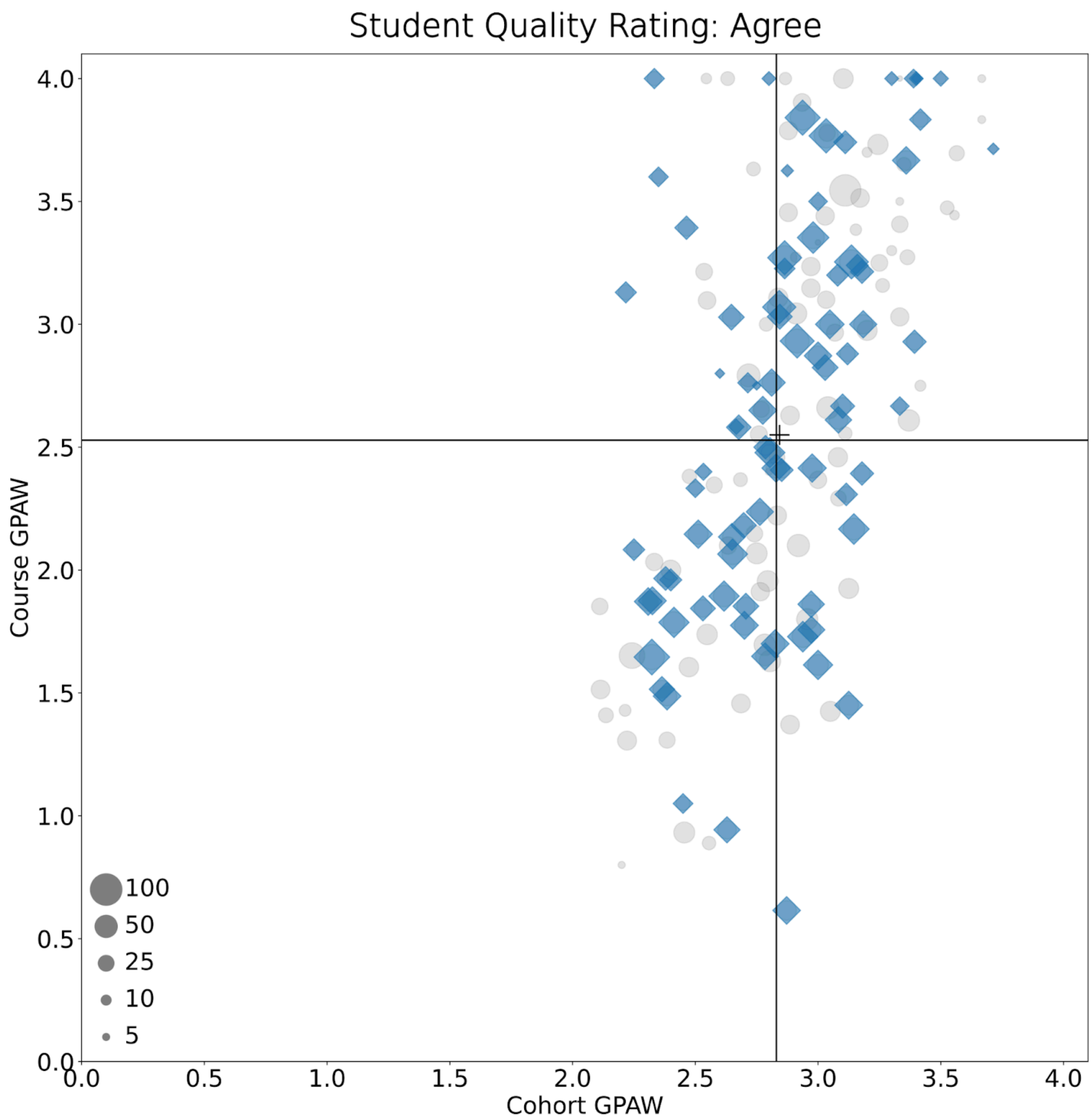

Figure 6.d. Faculty's Opinion of Student Quality - Agree Rating.

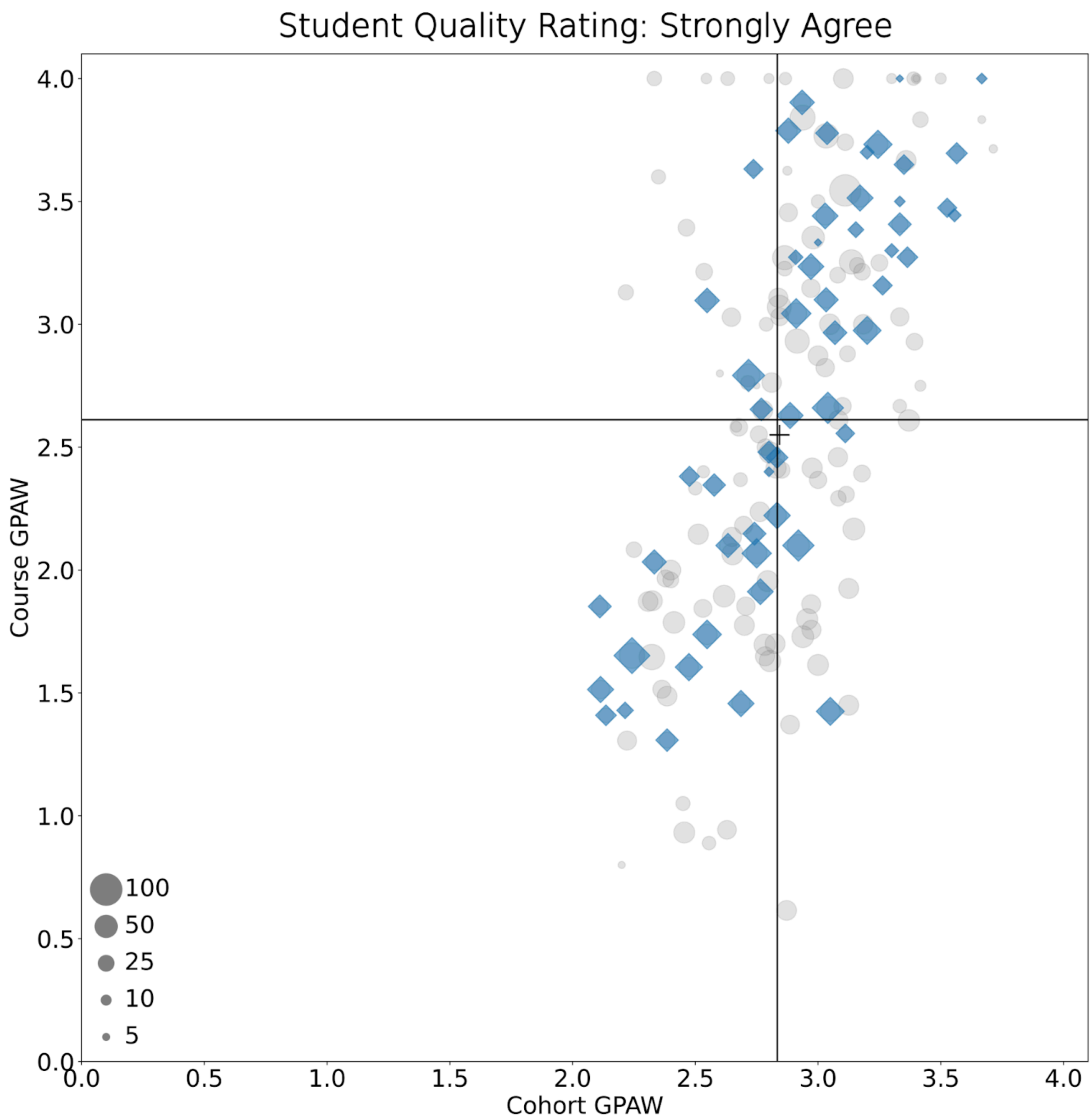

Figure 6.e. Faculty's Opinion of Student Quality - Strongly Agree Rating.

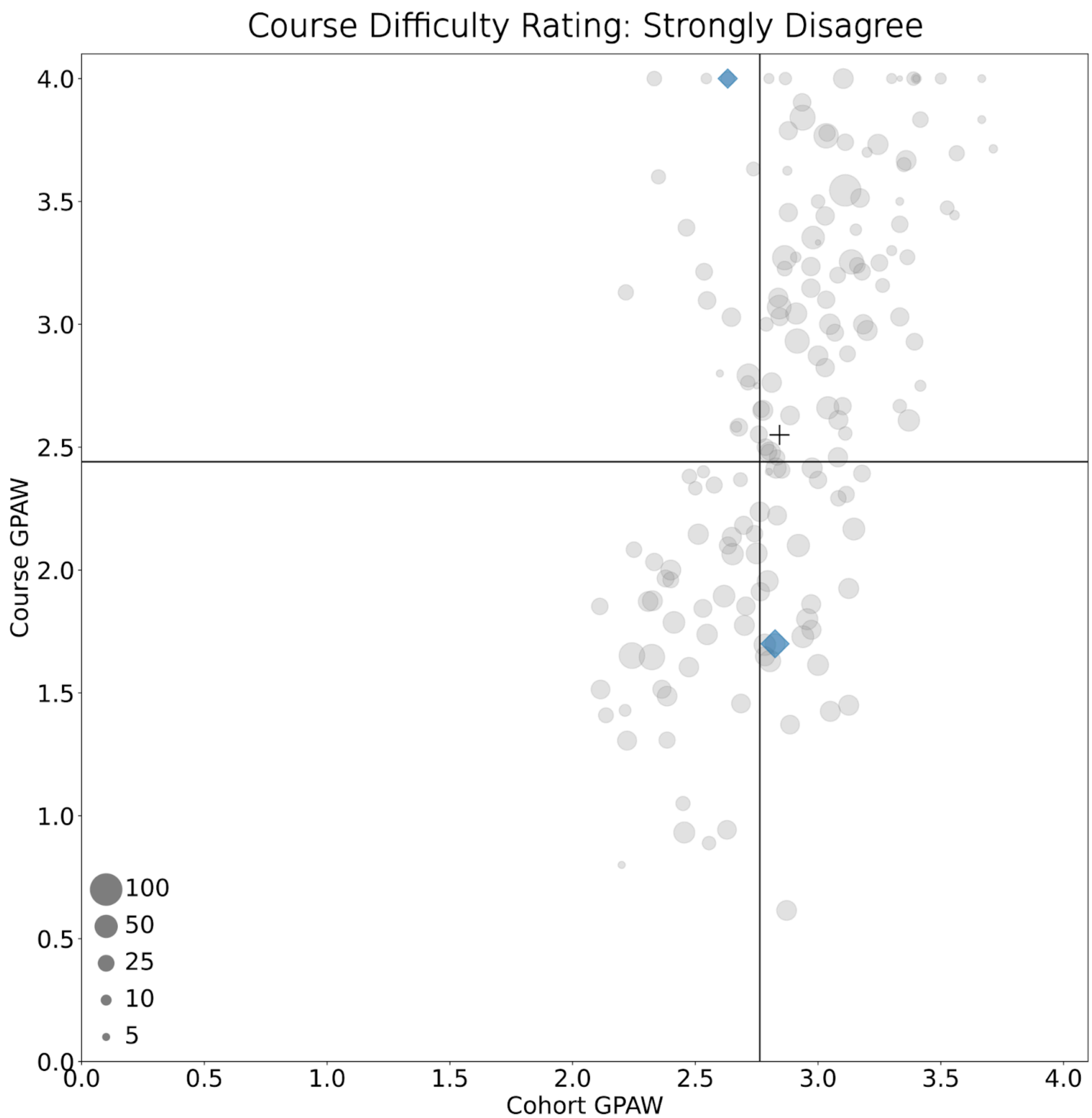

Figure 7.a. Faculty's Opinion of Course Difficulty - Strongly Disagree Rating.

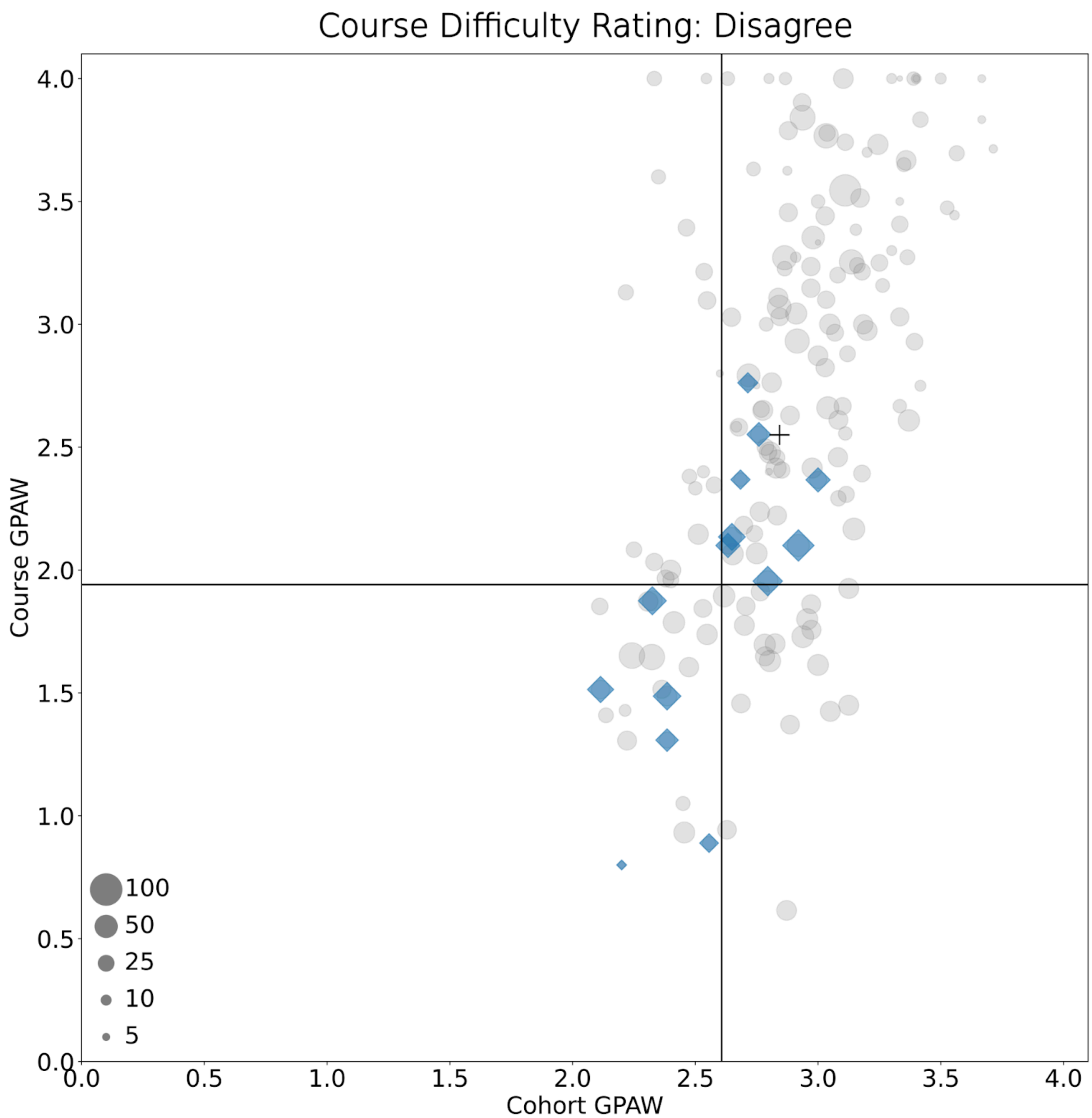

Figure 7.b. Faculty's Opinion of Course Difficulty - Disagree Rating.

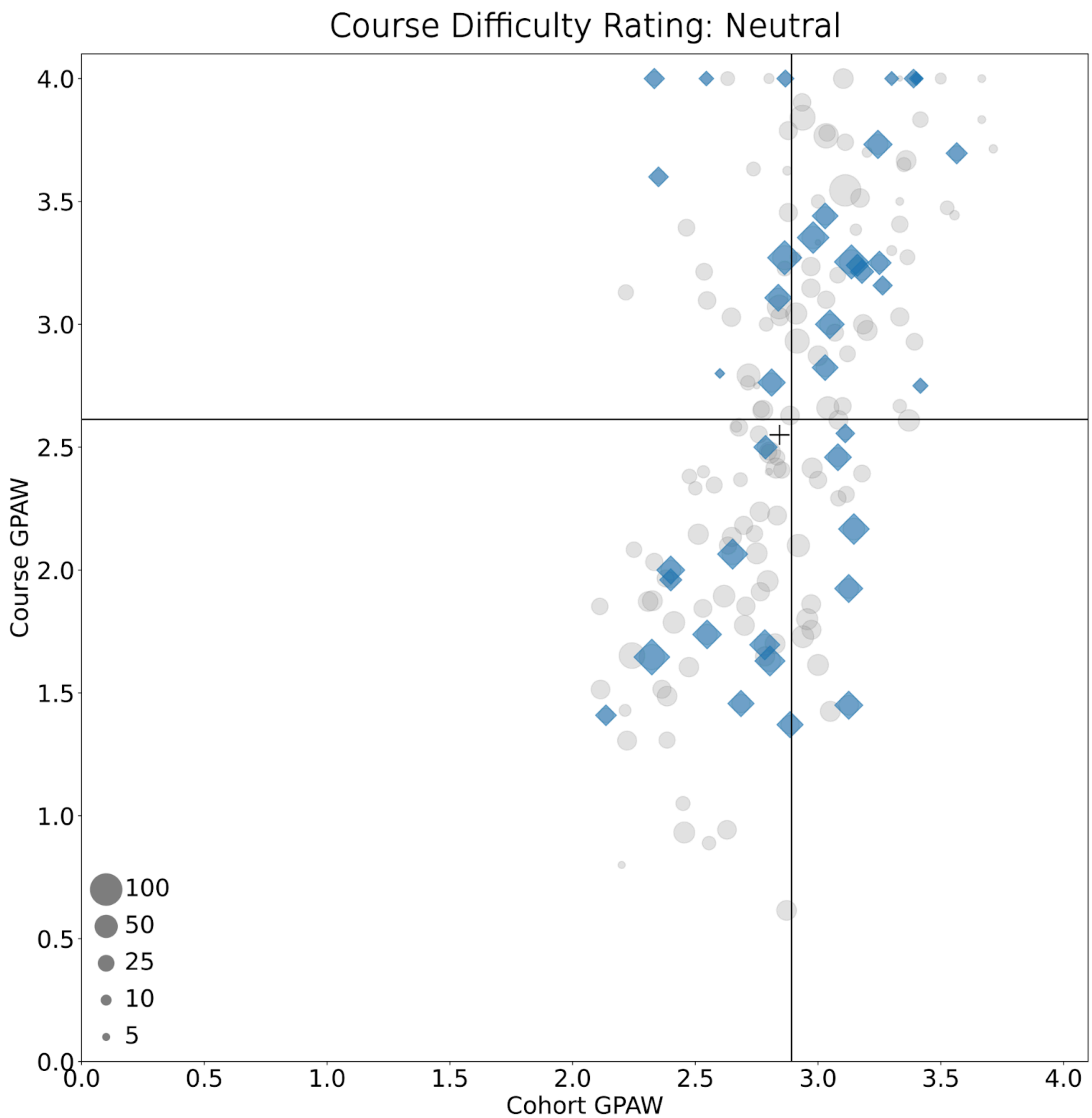

Figure 7.c. Faculty's Opinion of Course Difficulty - Neutral Rating.

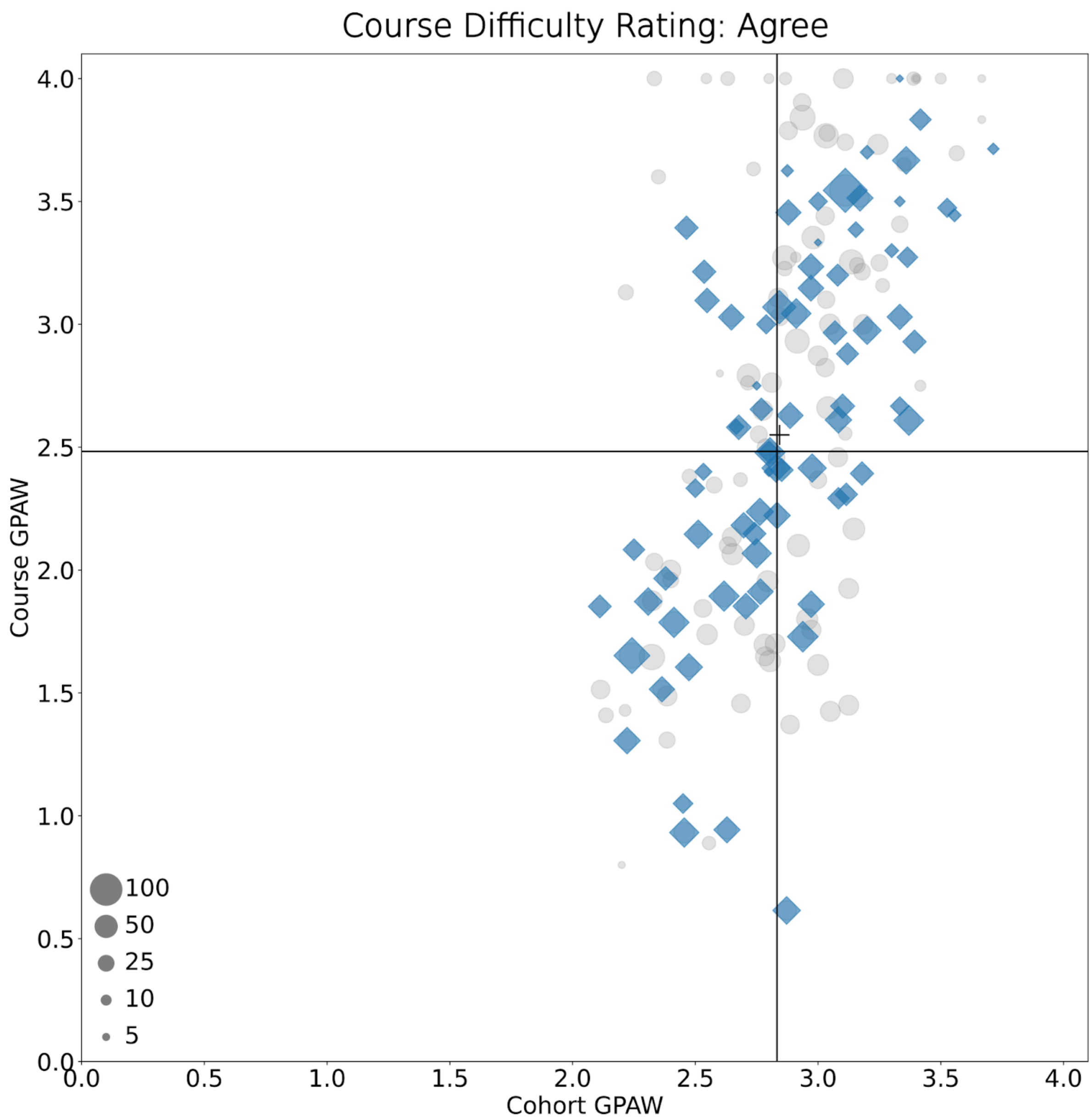

Figure 7.d. Faculty's Opinion of Course Difficulty - Agree Rating.

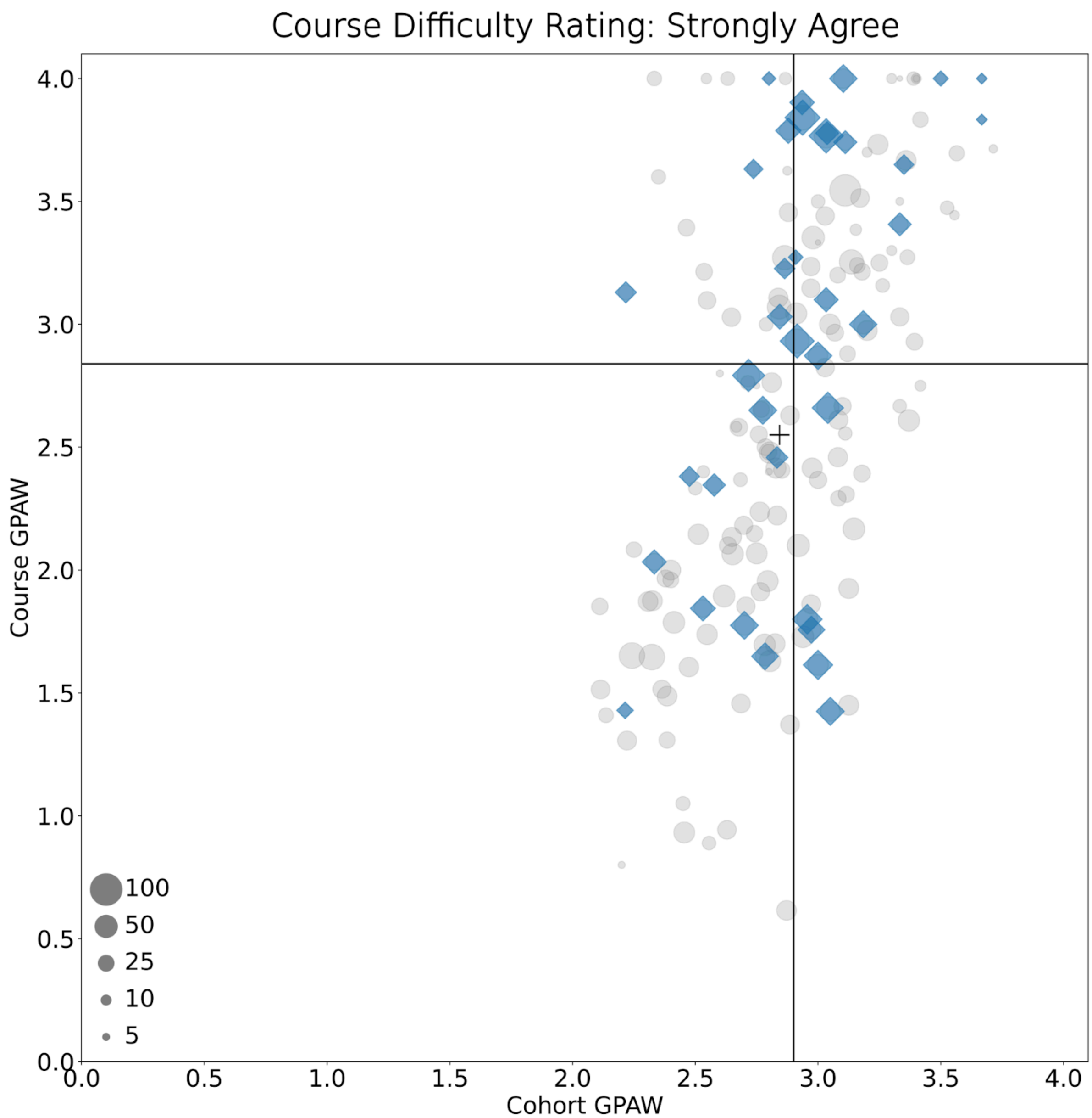

Figure 7.e. Faculty's Opinion of Course Difficulty - Strongly Agree Rating.

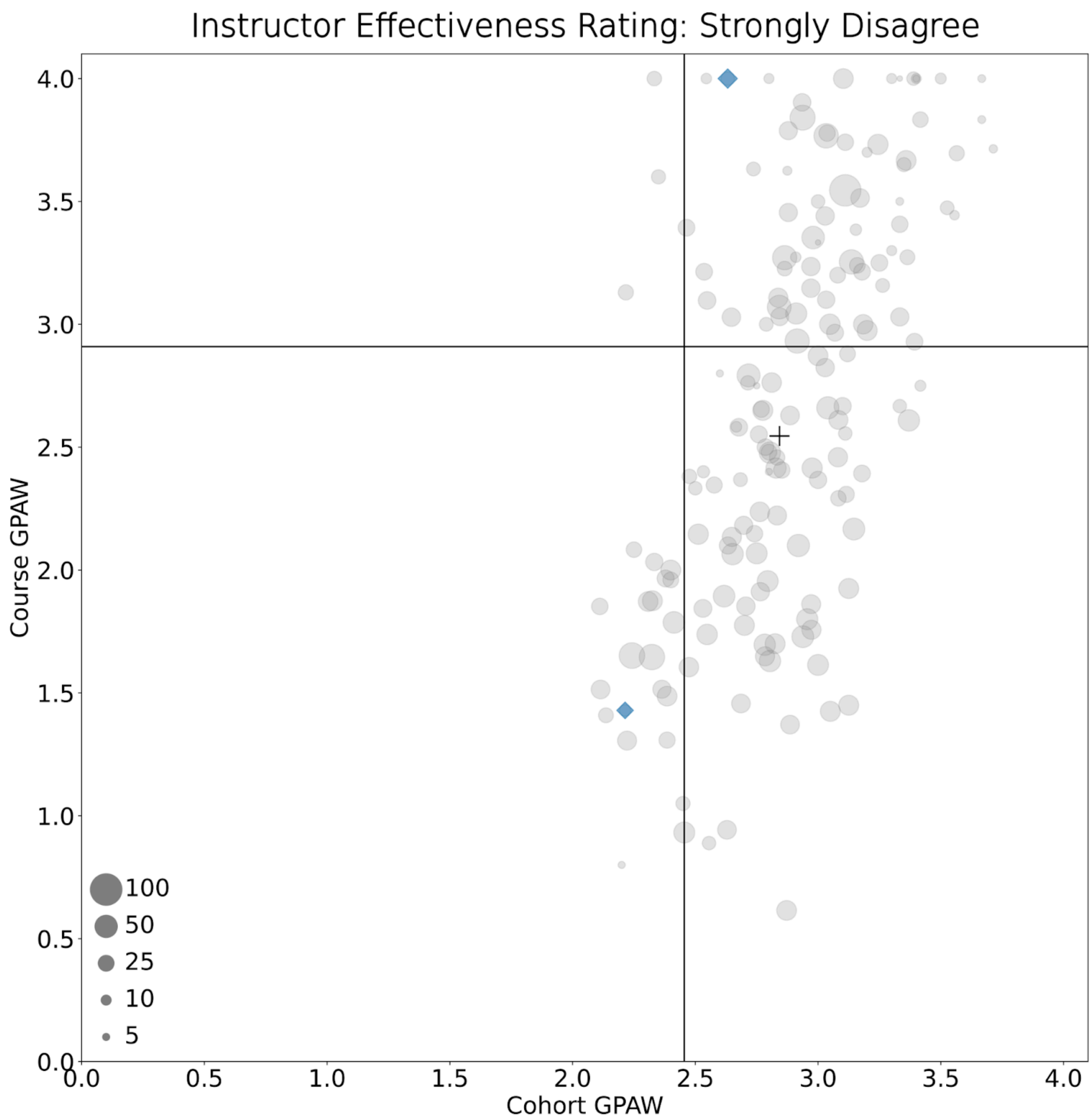

Figure 8.a. Faculty's Opinion of Instructor Effectiveness - Strongly Disagree Rating.

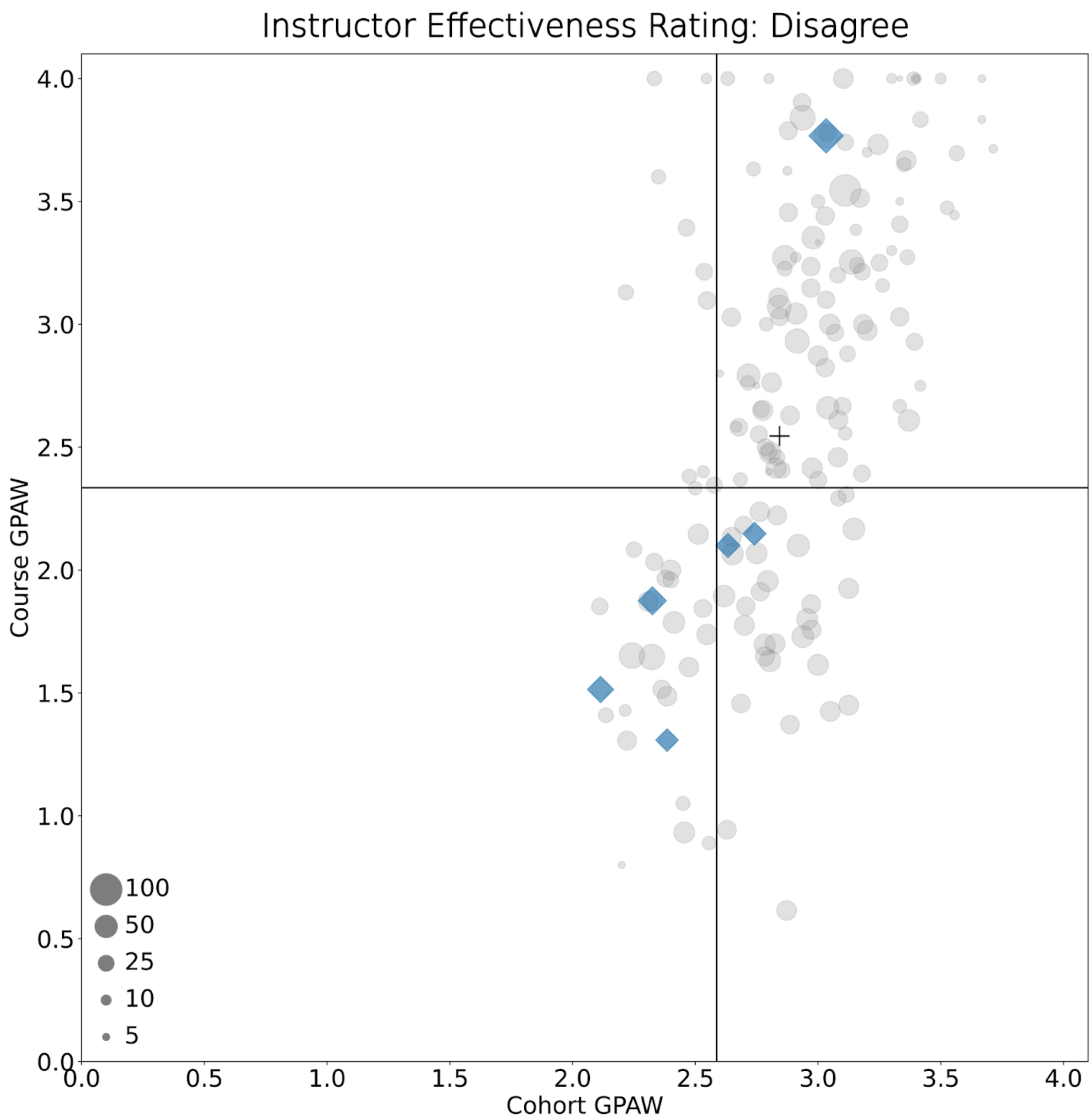

Figure 8.b. Faculty's Opinion of Instructor Effectiveness - Disagree Rating.

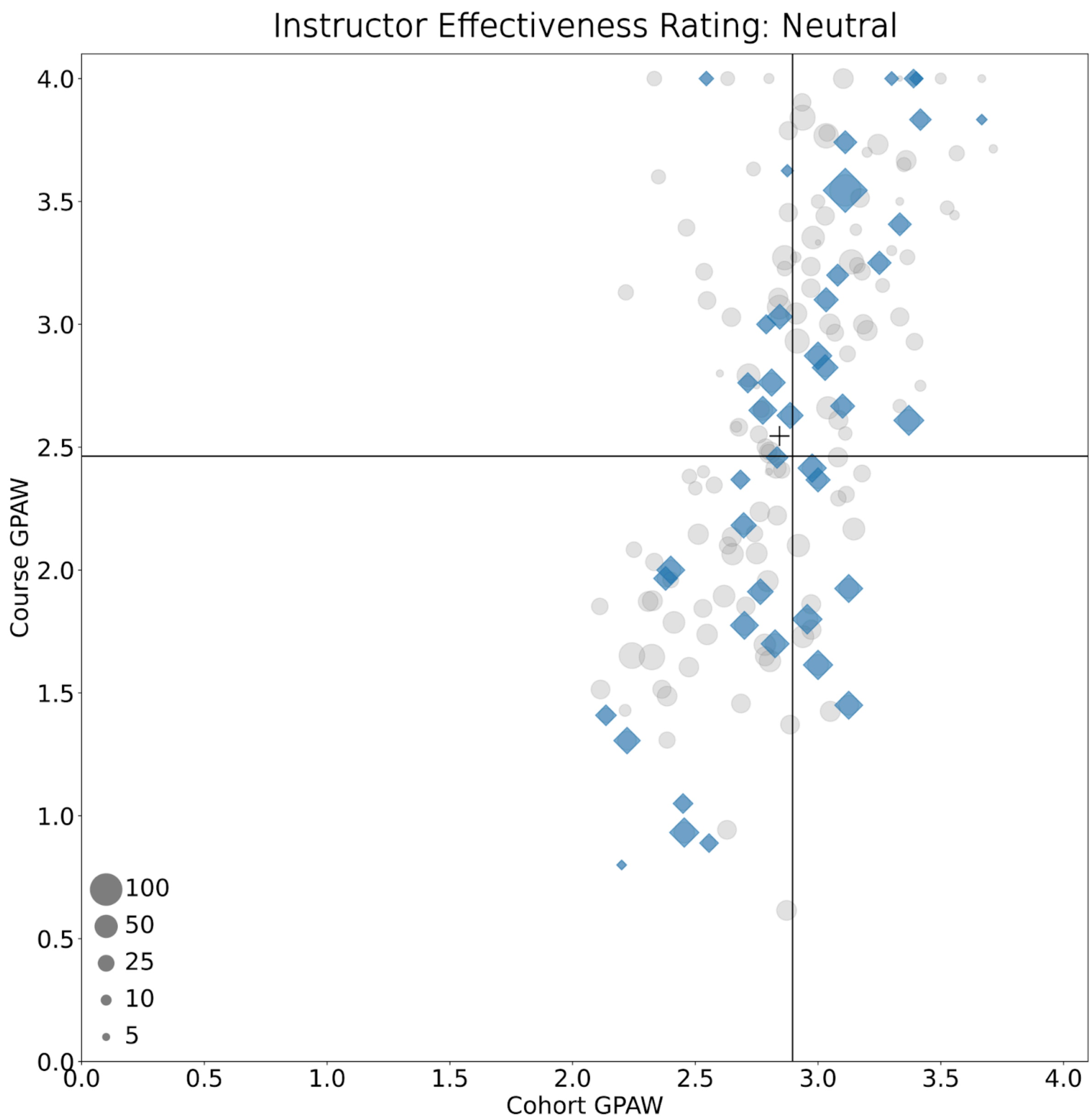

Figure 8.c. Faculty's Opinion of Instructor Effectiveness - Neutral Rating.

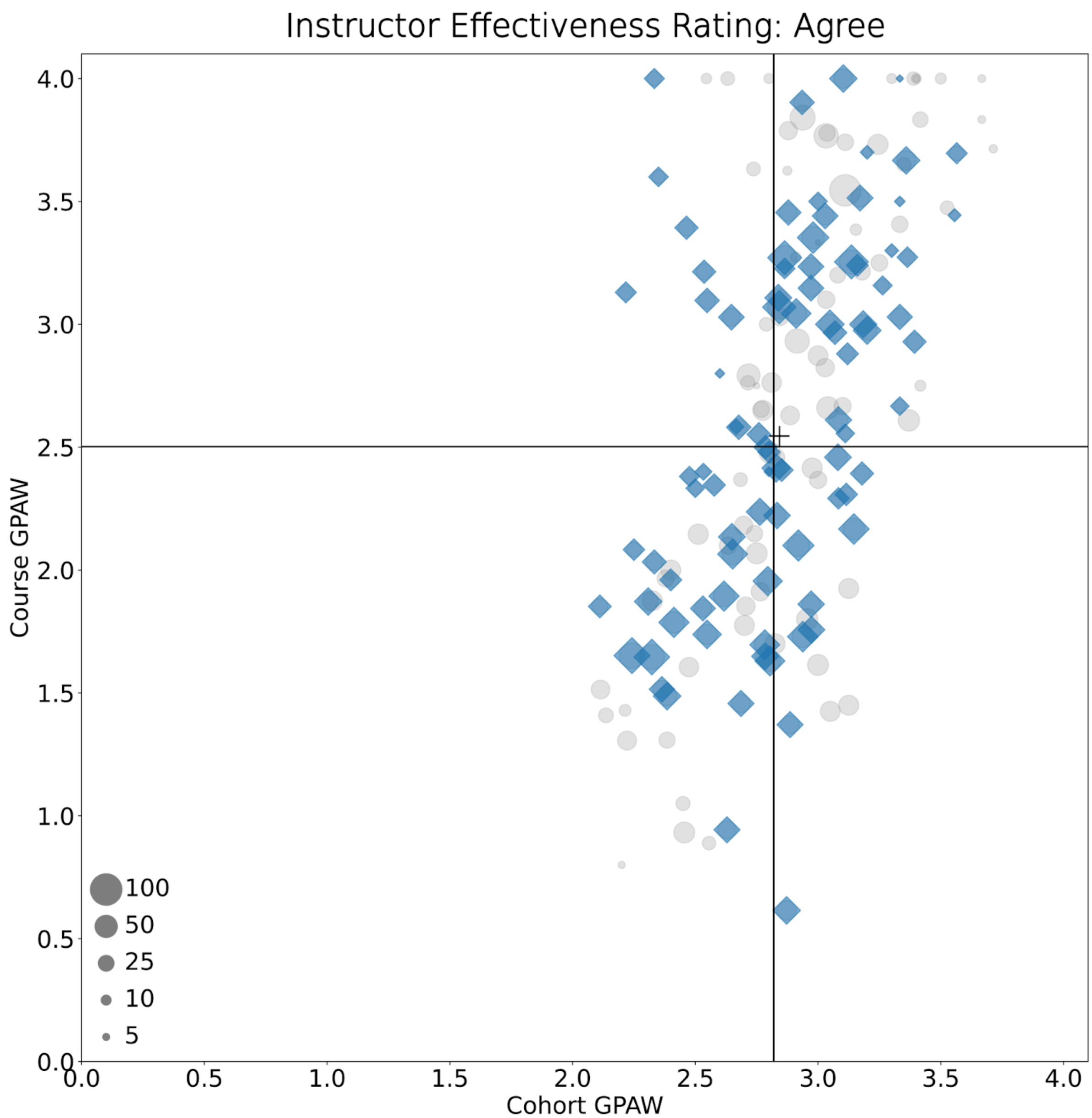

Figure 8.d. Faculty's Opinion of Instructor Effectiveness - Agree Rating.

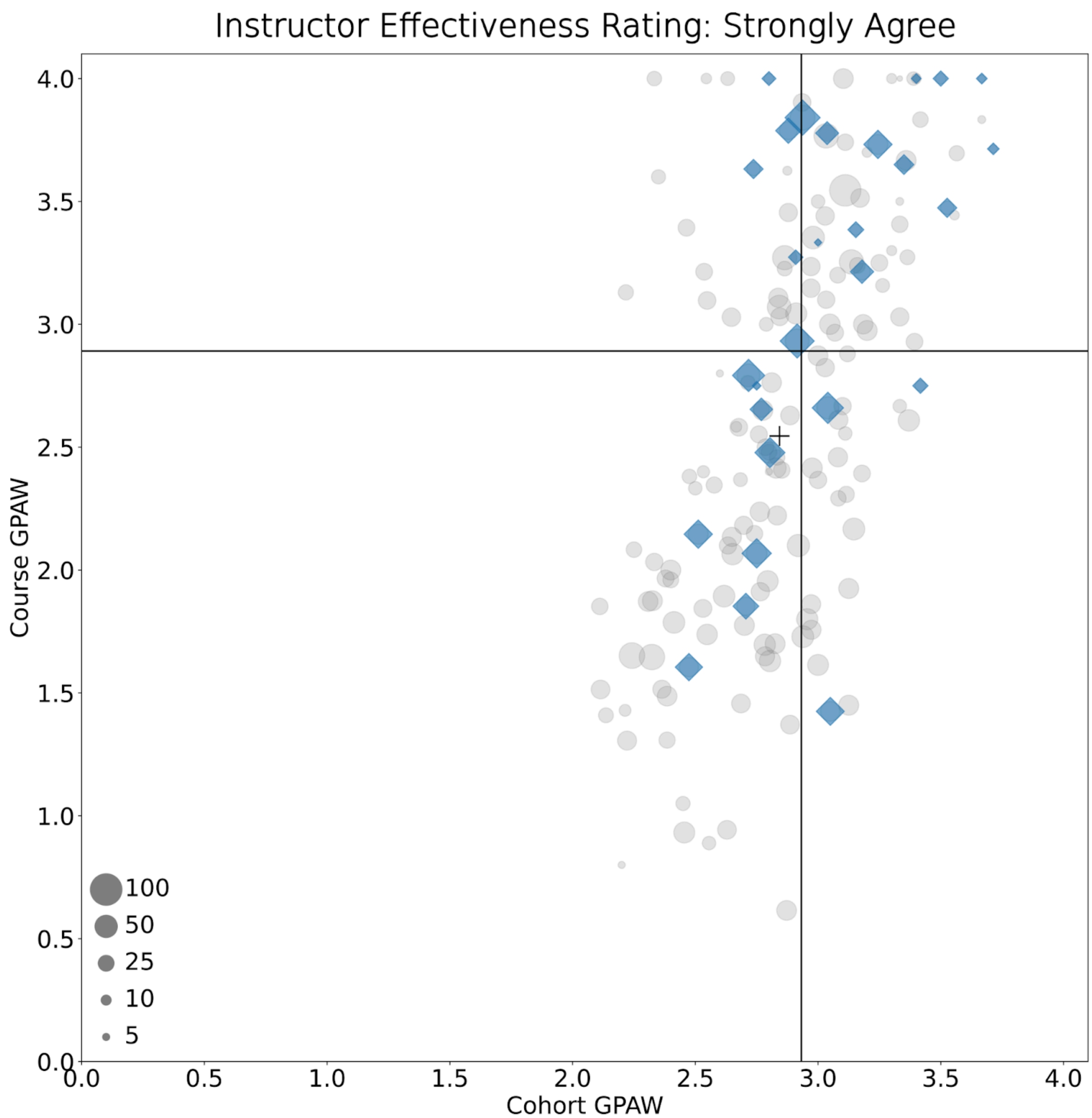

Figure 8.e. Faculty's Opinion of Instructor Effectiveness - Strongly Agree Rating.

References

[1] Manteufel, R. D., & Karimi, A. (2012, June), Assisting Adjunct Faculty Using Committee Exams Paper presented at 2012 ASEE Annual Conference & Exposition, San Antonio, Texas. 10.18260/1-2--20993

[2] Berk, Ronald A. "Survey of 12 strategies to measure teaching effectiveness." International journal of teaching and learning in higher education 17.1 (2005): 48-62.

[3] Medina, Melissa S., et al. "A review of strategies for designing, administering, and using student ratings of instruction." American journal of pharmaceutical education 83.5 (2019): 7177.

[4] Kacius, Carole, Cynthia Stone, and Silvia M. Bigatti. "Exit interviews: A decade of data to improve student learning experiences." (2015).

[5] Villanueva, Keisha A., et al. "Teaching evaluation practices in engineering programs: current approaches and usefulness." International Journal of Engineering Education 33.4 (2017): 1317-1334.

[6] Herman, Jennifer L., et al. "Creating a Peer Review of Teaching Process to Enhance Instructor Feedback in Engineering Education." 2021 ASEE Virtual Annual Conference Content Access. 2021.

[7] Simonson, Shawn R., Brittnee Earl, and Megan Frary. "Establishing a framework for assessing teaching effectiveness." College Teaching 70.2 (2022): 164-180.

[8] Godbout-Kinney, Keif, and Gavan PL Watson. "Institutional approaches to evaluate teaching effectiveness: The role of summative peer review of teaching for promotion and tenure." Canadian Journal of Educational Administration and Policy 201 (2022): 2-14.

[9] Quenette, Andrea, and Natalia Rybas. "Evaluating instructor quality: Moving away from student evaluations to peer review." New Directions for Teaching and Learning 2023.173 (2023): 67-74.

[10] Manteufel, Randall D., and Amir Karimi. "Assisting Adjunct Faculty Using Committee Exams." 2012 ASEE Annual Conference & Exposition. 2012.

[11] Hoyt, D.P., “Measurement of Instructional Effectiveness”, Research in Higher Education, Vol. 1, No. 4, 1973, pp. 367-378.