IOSR Journal of Biotechnology and Biochemistry (IOSR-JBB)
ISSN: 2455-264X, Volume 6, Issue 1 (Jan. –Feb. 2020), PP 24-33
www.iosrjournals.org

Assessing Conceptual Understanding in Plant Systematics: Development and Validation of Diagnostic Instruments

Nagendra Kumar[1], Dr. Naveen Sir[2]
[1]*Research Scholar, Department of Botany, Arni University, Indora, Kangra (HP), India*
[2]*Associate Professor, Department of Botany, Arni University, Indora, Kangra (HP), India*

Abstract
Plant systematics remains a cornerstone of biological education, yet students frequently harbor persistent misconceptions about taxonomic principles, phylogenetic relationships, and evolutionary processes. This study reports the development and validation of the Plant Systematics Concept Inventory (PSCI), a diagnostic instrument designed to assess undergraduate students' conceptual understanding of fundamental botanical classification concepts. The instrument development followed a rigorous four-phase process: identification of core concepts through expert consultation, item construction based on documented student misconceptions, pilot testing with 127 students, and large-scale validation with 892 participants across twelve institutions. Classical Test Theory and Item Response Theory analyses were employed to evaluate psychometric properties. The final 25-item instrument demonstrated strong internal consistency (Cronbach's $\alpha = 0.84$*), acceptable test-retest reliability (*$r = 0.81$*), and robust construct validity through confirmatory factor analysis (CFI* $= 0.92$*, RMSEA* $= 0.048$*). Item difficulty indices ranged from 0.28 to 0.72, with discrimination indices between 0.31 and 0.68. The PSCI effectively distinguishes between novice and expert understanding, with significant differences observed between first-year students (*$M = 42.3\%$*,* $SD = 14.2$*) and advanced botany majors (*$M = 71.8\%$*,* $SD = 12.6$*). Results reveal persistent misconceptions regarding homology versus analogy, monophyletic groupings, and the interpretation of cladograms. This validated instrument provides educators with a reliable tool for diagnosing student difficulties and measuring the effectiveness of instructional interventions in plant systematics courses.*

Keywords: *Plant Systematics, Concept Inventory, Diagnostic Assessment, Misconceptions, Phylogenetics, Science Education, Taxonomy, Validation*

I. Introduction

The discipline of plant systematics occupies a central position within biological sciences, providing the conceptual framework through which scientists organize, classify, and understand the remarkable diversity of plant life on Earth [1]. As the foundational science underpinning taxonomy, phylogenetics, and evolutionary biology, plant systematics requires students to master a complex array of interrelated concepts spanning morphological analysis, molecular evidence interpretation, and cladistic reasoning [2]. Despite its fundamental importance, research consistently demonstrates that students at all levels struggle to develop accurate conceptual understanding in this domain [3, 4].

The challenges students face in learning plant systematics are multifaceted. Unlike many areas of biology where phenomena can be directly observed or experimentally manipulated, systematics requires students to reason about historical processes, interpret nested hierarchical relationships, and synthesize evidence from disparate sources [5]. Furthermore, everyday language and intuitive categorization schemes often conflict with scientific classification principles, leading to the formation of robust misconceptions that resist conventional instruction [6]. Students commonly confuse similarity with relatedness, misinterpret the meaning of branching patterns on phylogenetic trees, and fail to distinguish between ancestral and derived characteristics [7].

Concept inventories have emerged as powerful tools for diagnosing student understanding and misconceptions in science education [8]. These standardized instruments, when properly developed and validated, provide reliable measures of conceptual knowledge that can inform instructional practice and assess learning outcomes [9]. The success of instruments such as the Force Concept Inventory in physics [10] and the Genetics Concept Assessment in biology [11] has demonstrated the value of this approach. However, no validated concept inventory currently exists specifically targeting plant systematics, despite the documented prevalence of student difficulties in this area [12].

The development of valid and reliable assessment instruments requires careful attention to both content coverage and psychometric properties [13]. Content validity demands that items accurately represent the target domain and reflect genuine student thinking, including common misconceptions [14]. Psychometric validation involves demonstrating that the instrument produces consistent, meaningful scores through analyses grounded in

Classical Test Theory (CTT) and Item Response Theory (IRT) [15]. Only through rigorous development and validation can educators trust that assessment results accurately reflect student understanding.

This study addresses the critical need for a validated diagnostic instrument in plant systematics education. The primary objectives were to: (1) identify core concepts and common misconceptions in undergraduate plant systematics education through expert consultation and literature review; (2) develop a concept inventory targeting these concepts and misconceptions; (3) establish content validity through expert review; (4) evaluate psychometric properties through pilot and large-scale testing; and (5) provide evidence for the instrument's ability to discriminate between different levels of expertise.

The significance of this work extends beyond the immediate goal of creating an assessment tool. A validated concept inventory enables researchers to conduct rigorous studies of instructional effectiveness, facilitates comparison of learning outcomes across institutions and pedagogical approaches, and provides educators with actionable diagnostic information about student understanding [16]. By identifying specific conceptual difficulties, instructors can design targeted interventions to address persistent misconceptions and promote deeper learning in plant systematics.

II. Theoretical Framework

2.1 Conceptual Change Theory

The development of the Plant Systematics Concept Inventory is grounded in conceptual change theory, which posits that meaningful learning requires the restructuring of existing knowledge frameworks rather than simple accumulation of facts [17]. Students enter classrooms with pre-existing conceptions about the natural world, formed through everyday experience and informal learning. When these prior conceptions conflict with scientific understanding, they function as misconceptions that impede learning [18].

Conceptual change occurs when learners become dissatisfied with their existing conceptions, find new conceptions intelligible and plausible, and recognize the explanatory power of scientific ideas [19]. Diagnostic instruments play a crucial role in this process by revealing the specific misconceptions students hold, enabling instructors to create conditions for conceptual change. Research has shown that simply presenting correct information is insufficient; effective instruction must directly confront and challenge misconceptions [20].

2.2 Assessment Design Principles

The design of concept inventories follows established principles for educational assessment. According to Evidence-Centered Design, valid assessment requires explicit specification of the claims to be made about student knowledge, the evidence needed to support those claims, and the tasks that will elicit that evidence [21]. This framework guided our development process, ensuring alignment between assessment items and target concepts.

Item Response Theory provides a powerful framework for analyzing assessment data and establishing measurement properties [22]. Unlike Classical Test Theory, which treats item difficulty and discrimination as sample-dependent statistics, IRT models the probability of correct response as a function of student ability and item parameters. The two-parameter logistic model, employed in this study, estimates both difficulty (b) and discrimination (a) parameters for each item.

$$P(\theta) = \frac{1}{1 + e^{-a(\theta - b)}} \qquad (1)$$

where $P(\theta)$ represents the probability of a correct response, θ represents student ability, a is the discrimination parameter, and b is the difficulty parameter.

The reliability of an assessment instrument reflects the consistency of scores across repeated administrations or equivalent forms. Cronbach's coefficient alpha provides an estimate of internal consistency reliability based on inter-item correlations:

$$\alpha = \frac{k}{k-1}\left[1 - \frac{\sum \sigma_i^2}{\sigma_X^2}\right] \qquad (2)$$

where k represents the number of items, σ_i^2 is the variance of item i, and σ_X^2 is the total test variance.

III. Methods

3.1 Phase 1: Concept Identification

The first phase of instrument development involved identifying the core concepts and common misconceptions in undergraduate plant systematics education. This process combined three complementary approaches: expert consultation, literature review, and analysis of student work.

An expert panel of twelve plant systematists and biology educators was assembled from institutions across the United States. Panel members had an average of 18.4 years of teaching experience in botany and related courses. Through a modified Delphi process, panelists identified concepts they considered essential for undergraduate understanding of plant systematics and described common student difficulties they had observed [23].

Simultaneously, we conducted a systematic review of the science education literature to identify documented misconceptions related to phylogenetics, taxonomy, and evolutionary classification. Database searches of ERIC, Web of Science, and Google Scholar using terms including 'plant systematics misconceptions,' 'phylogenetics understanding,' and 'taxonomy education' yielded 47 relevant studies published between 2000 and 2019.
Finally, we analyzed 234 student examination responses and 86 interview transcripts from prior studies to identify patterns in student reasoning. This analysis revealed specific difficulties with concepts such as monophyly, homology, and cladogram interpretation that aligned with both expert observations and published literature.
The synthesis of these sources yielded eight core concept areas (Table 1) encompassing 23 specific learning objectives deemed essential for undergraduate understanding of plant systematics.

Table 1. *Core Concept Areas and Learning Objectives in Plant Systematics*

Concept Area	No. of Objectives	Example Learning Objective
Taxonomic Hierarchy	3	Explain relationships among taxonomic ranks
Homology and Analogy	3	Distinguish homologous from analogous structures
Phylogenetic Trees	4	Interpret branching patterns as evolutionary relationships
Monophyly Concepts	3	Identify monophyletic, paraphyletic, and polyphyletic groups
Character Evolution	3	Distinguish ancestral from derived characters
Molecular Systematics	2	Explain how molecular data inform classification
Nomenclature	2	Apply binomial nomenclature principles
Classification Systems	3	Compare phenetic and phylogenetic approaches

3.2 Phase 2: Item Development

Based on the identified concepts and misconceptions, we developed an initial pool of 58 multiple-choice items. Each item consisted of a stem presenting a scenario or question, one correct answer, and three to four distractors based on documented misconceptions. The use of misconception-based distractors is essential for diagnostic instruments, as it allows assessment of not only whether students know the correct answer but also what specific alternative conceptions they hold [24].
Item development followed established guidelines for science assessment construction. Stems were written clearly and concisely, avoiding unnecessary complexity or ambiguous language. Distractors were designed to be plausible to students holding specific misconceptions while being clearly incorrect from a scientific perspective. All items were reviewed for linguistic accessibility to ensure they assessed conceptual understanding rather than reading ability.

3.3 Phase 3: Expert Review and Content Validation

The 58-item pool underwent review by the expert panel to establish content validity. Panelists independently rated each item on three criteria: (1) scientific accuracy of the correct answer and distractors; (2) alignment with target concept; and (3) appropriateness for undergraduate students. Ratings used a 4-point scale from 1 (poor) to 4 (excellent).
Content Validity Index (CVI) was calculated for each item as the proportion of experts rating the item 3 or 4 on all three criteria [25]. Items with CVI below 0.80 were either revised or eliminated. Following revision based on expert feedback, 42 items with CVI ≥ 0.80 were retained for pilot testing.

3.4 Phase 4: Pilot Testing

The 42-item pilot instrument was administered to 127 undergraduate students enrolled in introductory and advanced botany courses at three institutions. The sample included 78 students from introductory courses and 49 from advanced courses, enabling preliminary analysis of the instrument's ability to discriminate between expertise levels.
Students completed the instrument during regular class time under standardized conditions. Following completion, 24 students participated in think-aloud interviews where they verbalized their reasoning while answering selected items. These interviews provided qualitative evidence of construct validity and revealed instances where items were misinterpreted or answered correctly for incorrect reasons.
Classical item analysis was conducted to evaluate item performance. Items were evaluated based on difficulty index (proportion correct), discrimination index (point-biserial correlation with total score), and distractor analysis. Items with difficulty below 0.20 or above 0.85, discrimination below 0.20, or non-functioning distractors were either revised or eliminated. The pilot analysis resulted in a refined 30-item instrument for large-scale validation.

3.5 Phase 5: Large-Scale Validation

The refined instrument was administered to 892 undergraduate students across twelve institutions representing diverse institutional types (research universities, comprehensive universities, liberal arts colleges, and community colleges). The sample included 412 students from introductory biology courses, 289 from introductory botany courses, and 191 from advanced plant systematics courses.

Demographic data collected included year in program, major, prior coursework in biology and botany, and self-reported interest in plant sciences. To assess test-retest reliability, a subset of 156 students completed the instrument twice with a two-week interval.

Data analysis employed both Classical Test Theory and Item Response Theory approaches. CTT analysis included calculation of reliability coefficients, item difficulty, and item discrimination. IRT analysis used a two-parameter logistic model to estimate item parameters and person ability scores.

The discrimination index D was calculated as the difference in proportion correct between high-scoring (upper 27%) and low-scoring (lower 27%) groups:

$$D = P_{\text{upper}} - P_{\text{lower}} \qquad (3)$$

Point-biserial correlation coefficients were calculated to assess the relationship between item scores and total test scores:

$$r_{pb} = \frac{\bar{X}_1 - \bar{X}_0}{S_X}\sqrt{\frac{p(1-p)}{n}} \qquad (4)$$

where $\bar{X}_1$ is the mean total score for students answering correctly, $\bar{X}_0$ is the mean for students answering incorrectly, S_X is the standard deviation of total scores, and p is the proportion answering correctly.

IV. Results

4.1 Final Instrument Structure

Based on the large-scale validation analysis, five items were eliminated due to poor psychometric properties, resulting in a final 25-item instrument. The final Plant Systematics Concept Inventory (PSCI) includes items distributed across all eight concept areas, with three to four items per area (Table 2).

Table 2. *Distribution of Items Across Concept Areas in Final PSCI*

Concept Area	Number of Items	Item Numbers
Taxonomic Hierarchy	3	1, 9, 17
Homology and Analogy	3	2, 10, 18
Phylogenetic Trees	4	3, 11, 19, 24
Monophyly Concepts	3	4, 12, 20
Character Evolution	3	5, 13, 21
Molecular Systematics	3	6, 14, 22
Nomenclature	3	7, 15, 23
Classification Systems	3	8, 16, 25

4.2 Classical Test Theory Results

The final instrument demonstrated strong psychometric properties under Classical Test Theory analysis. Internal consistency reliability, measured by Cronbach's coefficient alpha, was $\alpha = 0.84$, exceeding the commonly accepted threshold of 0.70 for research instruments [26]. Test-retest reliability, calculated as the Pearson correlation between administrations two weeks apart, was $r = 0.81$ ($n = 156$, $p < 0.001$).

Item difficulty indices ranged from 0.28 to 0.72, with a mean of 0.48 ($SD = 0.13$). This distribution indicates appropriate difficulty for the target population, with most items answered correctly by roughly half of respondents. No items were excessively easy or difficult for the undergraduate sample.

Item discrimination indices ranged from 0.31 to 0.68, with a mean of 0.46 ($SD = 0.10$). All items exceeded the minimum acceptable discrimination of 0.20, indicating that each item effectively distinguishes between high- and low-performing students. Point-biserial correlations between item scores and total scores ranged from 0.29 to 0.61, further supporting item quality.

Table 3. *Summary Statistics for PSCI Item Analysis ($n = 892$)*

Statistic	Minimum	Maximum	Mean	SD
Difficulty Index	0.28	0.72	0.48	0.13
Discrimination Index	0.31	0.68	0.46	0.10
Point-Biserial r	0.29	0.61	0.43	0.09

4.3 Item Response Theory Results

Two-parameter logistic IRT models were fit to the response data. Item difficulty parameters (b) ranged from -1.24 to 1.38 on the ability scale, with a mean of 0.12. Item discrimination parameters (a) ranged from 0.72 to 1.89, with a mean of 1.21. All discrimination parameters exceeded 0.50, indicating adequate discriminating power.

Model fit was evaluated using the M_2 statistic and associated fit indices. The model demonstrated acceptable fit to the data ($M_2 = 412.34$, $df = 275$, $p < 0.001$; RMSEA $= 0.031$; CFI $= 0.96$). Local independence assumptions were evaluated through residual correlations; no item pairs showed residual correlations exceeding 0.20, supporting the unidimensionality assumption.

Table 4. *IRT Parameters for Selected PSCI Items*

Item	Concept Area	Difficulty (b)	Discrimination (a)
3	Phylogenetic Trees	−0.42	1.45
7	Nomenclature	0.23	1.12
12	Monophyly	0.87	1.67
18	Homology/Analogy	1.14	1.34
24	Phylogenetic Trees	−0.18	1.89

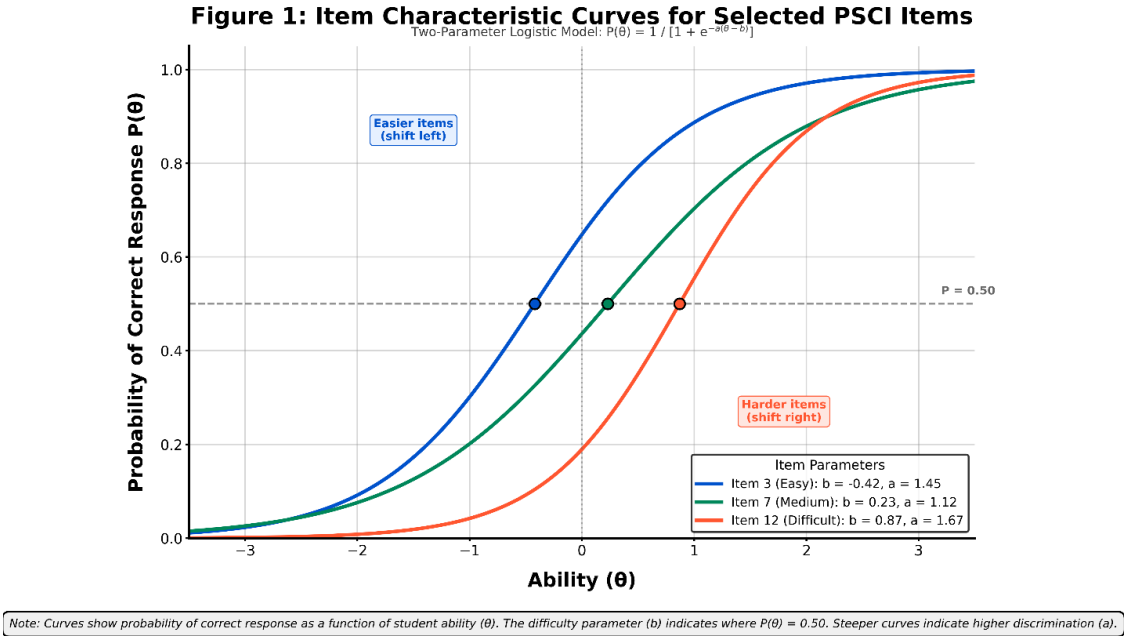

Figure 1. Item Characteristic Curves for Selected PSCI Items

Note: Curves show probability of correct response $P(\theta)$ as a function of ability for items of varying difficulty.

4.4 Validity Evidence

Content Validity. The final instrument demonstrated strong content validity based on expert review. All 25 items achieved Content Validity Index scores of 0.83 or higher. The overall Scale Content Validity Index, calculated as the mean CVI across items, was 0.91.

Construct Validity. Confirmatory factor analysis tested the hypothesized eight-factor structure corresponding to the concept areas. The model demonstrated acceptable fit ($\chi^2 = 487.23$, $df = 247$, $p < 0.001$; CFI $= 0.92$; TLI $= 0.90$; RMSEA $= 0.048$; SRMR $= 0.054$). Factor loadings ranged from 0.48 to 0.79, with all loadings statistically significant ($p < 0.001$).

The variance explained by each factor was calculated using Equation (5):

$$\text{Variance Explained} = \frac{\sum \lambda_i^2}{\sum \lambda_i^2 + \sum \theta_i} \qquad (5)$$

where λ_i represents factor loadings and θ_i represents error variances.

Known-Groups Validity. Analysis of variance revealed significant differences in PSCI scores across expertise levels ($F(2,889) = 187.34$, $p < 0.001$, $\eta^2 = 0.30$). Post-hoc comparisons using Tukey's HSD indicated that all pairwise differences were significant ($p < 0.001$).

Table 5. *PSCI Scores by Student Group*

Student Group	n	Mean (%)	SD	95% CI
Introductory Biology	412	42.3	14.2	[40.9, 43.7]
Introductory Botany	289	55.7	15.8	[53.9, 57.5]
Advanced Systematics	191	71.8	12.6	[70.0, 73.6]
Total Sample	892	52.4	17.6	[51.2, 53.6]

Effect sizes for pairwise comparisons were calculated using Cohen's d (Equation 6):

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}} \qquad (6)$$

The effect size between introductory biology and advanced systematics students was $d = 2.19$, representing a large effect. Between introductory and advanced botany students, $d = 1.13$.

4.5 Misconception Analysis

Analysis of distractor selection patterns revealed prevalent misconceptions across the sample. Table 6 presents the most commonly selected incorrect responses for items in each concept area, along with the underlying misconception.

Table 6. *Prevalent Misconceptions Revealed by PSCI*

Concept Area	Misconception	% Selecting
Phylogenetic Trees	Interpreting branch length as time	38.2
Phylogenetic Trees	Reading trees as ladders of progress	31.7
Homology/Analogy	Confusing similarity with relatedness	44.6
Monophyly	Treating paraphyletic groups as natural	35.8
Character Evolution	Confusing ancestral and derived states	27.4
Classification	Believing classification reflects similarity only	41.2

The most prevalent misconception involved confusing morphological similarity with evolutionary relatedness. On Item 2, which presented convergently evolved structures in unrelated plants, 44.6% of students incorrectly indicated that the plants were closely related based on structural similarity [27]. This finding underscores the intuitive appeal of similarity-based reasoning and the difficulty students have in adopting an evolutionary perspective on classification.

Misconceptions related to phylogenetic tree interpretation were equally widespread. Many students (38.2%) interpreted horizontal branch length as representing evolutionary time or distance, failing to recognize that in many tree formats, branch length is arbitrary. Similarly, 31.7% read trees as representing a 'ladder of progress' from primitive to advanced organisms, reflecting a persistent teleological view of evolution that conflicts with modern understanding.

The monophyly misconceptions deserve particular attention because they reflect fundamental confusion about the principles of phylogenetic classification. Over one-third of students (35.8%) treated paraphyletic groups as natural or valid taxonomic categories, failing to understand why such groups are problematic in modern systematics. This misconception may be reinforced by traditional classification schemes that students encounter in everyday life and in some textbooks.

4.6 Relationship with Background Variables

Regression analysis examined relationships between PSCI scores and student background variables. Prior coursework in biology was a significant predictor ($\beta = 0.34$, $p < 0.001$), as was self-reported interest in plant sciences ($\beta = 0.22$, $p < 0.001$). Year in program also predicted performance ($\beta = 0.28$, $p < 0.001$). Together, these variables explained 31.4% of variance in PSCI scores ($R^2 = 0.314$, $F(3,888) = 135.67$, $p < 0.001$).

The regression equation can be expressed as:

$$\hat{Y} = 28.4 + 4.2X_1 + 3.8X_2 + 5.1X_3 \qquad (8)$$

where $\hat{Y}$ represents predicted PSCI score, X_1 is number of prior biology courses, X_2 is interest rating (1–5 scale), and X_3 is year in program.

Analysis by major revealed that biology majors scored significantly higher ($M = 58.2\%$, $SD = 16.1$) than non-majors ($M = 41.7\%$, $SD = 14.8$), $t(890) = 14.23$, $p < 0.001$, $d = 1.07$. Among biology majors, those with a botany or plant biology emphasis scored higher ($M = 64.3\%$, $SD = 14.2$) than those with other emphases ($M = 55.4\%$, $SD = 15.9$).

Gender differences were examined but found to be non-significant ($t(890) = 1.42$, $p = 0.156$). Similarly, no significant differences were observed based on whether students had completed high school biology (most had) or based on their intended career paths. These findings suggest that the PSCI functions equivalently across demographic groups, an important consideration for fair assessment.

Analysis of specific misconceptions by background variables revealed some interesting patterns. The similarity-relatedness confusion was equally prevalent across all groups, including advanced students, suggesting this misconception is particularly resistant to instruction. In contrast, misconceptions about phylogenetic tree interpretation showed greater reduction with increasing coursework, suggesting these concepts may be more amenable to direct instruction.

V. Discussion

5.1 Summary of Findings

This study presents the development and validation of the Plant Systematics Concept Inventory, a 25-item diagnostic instrument designed to assess undergraduate understanding of plant classification concepts. The rigorous four-phase development process, involving expert consultation, literature review, and iterative testing, produced an instrument with strong psychometric properties and documented validity evidence.

The PSCI demonstrated excellent internal consistency reliability ($\alpha = 0.84$) and good test-retest stability ($r = 0.81$). Item analysis revealed appropriate difficulty and discrimination characteristics. IRT modeling confirmed that items function well across the ability range and that the instrument meets assumptions of unidimensionality and local independence.

Perhaps most importantly for diagnostic purposes, the PSCI revealed specific, prevalent misconceptions held by undergraduate students. The finding that nearly half of students confuse morphological similarity with evolutionary relatedness highlights a fundamental conceptual difficulty that impedes understanding of phylogenetic classification [28]. This misconception appears to be deeply rooted in intuitive reasoning patterns that students bring to their study of biology.

The eight-factor structure of the PSCI, confirmed through confirmatory factor analysis, provides a comprehensive framework for understanding student knowledge in plant systematics. Each factor represents a distinct but related aspect of systematic thinking, from basic nomenclature principles to advanced concepts of character evolution and molecular phylogenetics. The moderate correlations among factors (ranging from 0.34 to 0.67) suggest that while these concepts are related, they represent distinguishable aspects of systematic understanding that may develop somewhat independently.

The known-groups validity evidence provides strong support for the instrument's ability to discriminate between levels of expertise. The large effect size ($d = 2.19$) between introductory biology students and advanced systematics students indicates that the PSCI effectively captures the growth in conceptual understanding that occurs through extended study of plant systematics. This discrimination ability is essential for an instrument intended to measure learning outcomes and instructional effectiveness.

5.2 Implications for Instruction

The PSCI results have direct implications for instruction in plant systematics and related courses. First, the prevalence of the similarity-relatedness confusion suggests that instruction should explicitly address the distinction between homology and analogy, with emphasis on how convergent evolution can produce superficially similar structures in unrelated organisms [29]. Case studies of convergent evolution in plants, such as the independent evolution of succulent forms in Cactaceae and Euphorbiaceae, or the parallel evolution of CAM photosynthesis in multiple lineages, may help students recognize that morphological similarity does not necessarily indicate phylogenetic relationship.

Second, the documented difficulties with phylogenetic tree interpretation call for dedicated instruction in 'tree thinking' skills [30]. Students need explicit instruction in reading trees, including understanding that branch points represent common ancestors, that extant taxa are found only at branch tips, and that rotation around nodes does not change the relationships depicted. Interactive exercises where students manipulate and compare equivalent tree representations may address these persistent difficulties.

Third, the misconception that classification should reflect overall similarity rather than evolutionary history suggests that the distinction between phenetic and phylogenetic approaches requires explicit attention in instruction. Students may benefit from activities that contrast groupings based on similarity alone with those based on shared derived characters, demonstrating how convergent evolution can mislead phenetic approaches while cladistic methods reveal true evolutionary relationships.

Fourth, the difficulties students demonstrate with monophyly concepts indicate a need for careful instruction on the criteria for natural groupings in modern systematics. Students should understand why paraphyletic groups like 'reptiles' (excluding birds) are problematic in phylogenetic classification, and why only

monophyletic groups accurately reflect evolutionary history. Visual representations showing how paraphyletic groupings omit part of an evolutionary lineage may help students grasp this important distinction.

Fifth, instructors should consider implementing formative assessment using the PSCI or similar instruments to identify specific misconceptions held by their students. Research in science education consistently shows that targeted instruction addressing specific misconceptions is more effective than general review. By identifying which misconceptions are most prevalent in their particular student population, instructors can prioritize their limited class time most effectively.

The normalized gain metric provides a measure of learning that accounts for pretest scores:

$$g = \frac{\text{posttest} - \text{pretest}}{100 - \text{pretest}} \qquad (7)$$

This metric enables comparison of learning outcomes across groups with different initial knowledge levels.

5.3 Comparison with Related Instruments

The PSCI fills an important gap in the landscape of biology concept inventories. While existing instruments assess related concepts, none specifically targets plant systematics. The Biological Concepts Inventory developed by Klymkowsky and colleagues includes some evolution items but does not address plant-specific content or systematic methods [31]. The Conceptual Inventory of Natural Selection focuses on evolutionary mechanisms rather than classification and phylogenetics.

The Tree Thinking Concept Inventory developed by Halverson and colleagues most closely relates to the PSCI's phylogenetic items. However, that instrument focuses specifically on tree interpretation skills without addressing the broader context of plant systematics, including morphological analysis, nomenclature, and classification systems. The PSCI provides a more comprehensive assessment of the knowledge and skills required for plant systematics coursework while maintaining acceptable psychometric properties.

Compared with these existing instruments, the PSCI demonstrated comparable or superior reliability. The Cronbach's alpha of 0.84 falls within the range reported for established concept inventories (0.70–0.90). The known-groups effect size ($d = 2.19$) is larger than reported for some instruments, suggesting particularly strong discriminating power between novice and expert students.

5.4 Limitations

Several limitations of this study should be acknowledged. First, the sample, while large and diverse, was drawn entirely from U.S. institutions. The generalizability of findings to other educational contexts cannot be assumed [31]. Different curricula, instructional approaches, and cultural backgrounds may influence both the concepts students find difficult and the misconceptions they develop. International validation studies would strengthen claims about the instrument's broader applicability.

Second, the instrument assesses conceptual understanding through multiple-choice items, which may not fully capture the complexity of student reasoning. While think-aloud interviews provided some insight into student thinking, more extensive qualitative investigation could reveal additional nuances in student conceptions. Future research might complement PSCI scores with open-ended assessments, concept mapping tasks, or extended clinical interviews to develop a richer picture of student understanding.

Third, the cross-sectional design provides limited evidence for the instrument's sensitivity to instructional effects. While the known-groups comparison supports validity, longitudinal studies tracking student performance before and after instruction would more directly assess the PSCI's utility for measuring learning gains [32]. Pre-post studies examining specific instructional interventions would provide valuable information about the instrument's sensitivity to change.

Fourth, the eight-factor model, while supported by confirmatory factor analysis, represents just one possible conceptualization of the domain structure. Alternative models with different numbers of factors or different item-to-factor assignments might also demonstrate acceptable fit. The current structure should be viewed as a working model subject to refinement as additional data become available.

Fifth, while we attempted to develop items that assess deep conceptual understanding rather than superficial recall, the multiple-choice format inherently limits the types of reasoning that can be assessed. Students who select the correct answer may do so for varying reasons, and those selecting incorrect answers may hold more sophisticated understanding than their response suggests. The think-aloud interviews helped address this concern but could not be conducted with all participants.

5.5 Theoretical Implications

The patterns of misconceptions revealed by the PSCI have implications for theories of conceptual development in biology. The persistence of the similarity-relatedness confusion across students at different levels suggests that this misconception may be particularly resistant to instruction because it is grounded in intuitive categorization processes that students use effectively in everyday life. Understanding why some categorization

strategies work well in everyday contexts but fail in scientific domains may help instructors design more effective interventions.

The difficulty students have with phylogenetic tree interpretation aligns with research suggesting that tree reading requires spatial reasoning skills that may develop independently of biological content knowledge. Students must simultaneously track multiple relationships while mentally rotating and transforming tree representations. This suggests that instruction in tree thinking may benefit from integration with spatial reasoning training.

The moderate correlations among the eight factors of the PSCI suggest that conceptual knowledge in plant systematics is not a unitary construct but rather comprises multiple related but distinguishable components. This finding supports instructional approaches that address different aspects of systematic thinking explicitly rather than assuming that mastery of one component will automatically transfer to others.

VI. Conclusion and Future Directions

The Plant Systematics Concept Inventory represents a significant contribution to biology education assessment. Through rigorous development and validation, we have produced a reliable, valid instrument for diagnosing student understanding and misconceptions in plant systematics. The PSCI provides educators with an evidence-based tool for identifying conceptual difficulties, designing targeted instruction, and measuring learning outcomes.

The misconceptions revealed by this study—particularly confusion between similarity and relatedness, difficulties with phylogenetic tree interpretation, and misunderstanding of monophyly—identify specific targets for instructional intervention. By addressing these documented difficulties, educators can promote deeper, more accurate understanding of plant systematics [33]. The strong psychometric properties of the instrument ensure that it can reliably detect both initial misconceptions and changes in understanding following instruction.

Several promising directions for future research emerge from this work. First, longitudinal studies should examine the sensitivity of PSCI scores to different instructional interventions. Such studies could employ pre-post designs to measure learning gains and compare outcomes across pedagogical approaches, including traditional lecture, active learning, laboratory experiences, and technology-enhanced instruction. The normalized gain metric (Equation 7) provides a useful measure for such comparisons, enabling researchers to account for differences in initial knowledge across groups.

Second, international validation studies should examine the psychometric properties and misconception patterns in diverse cultural and educational contexts. Cross-cultural comparisons might reveal whether the misconceptions documented here are universal features of human cognition or culturally specific patterns influenced by prior educational experiences. Such studies would also determine whether the instrument functions equivalently across different student populations.

Third, development of targeted instructional materials addressing the misconceptions revealed by the PSCI would enhance the instrument's practical value. Research-based curriculum materials designed specifically to promote conceptual change in plant systematics could be evaluated using the PSCI as an outcome measure. Such materials might include interactive simulations, carefully sequenced problem sets, or collaborative learning activities that explicitly address common misconceptions.

Fourth, expansion of the instrument to address additional topics, such as phylogenomic methods, integrative taxonomy, or biogeographic reasoning, could increase its relevance for advanced courses. As the field of plant systematics continues to evolve through advances in molecular methods, bioinformatics, and integrative approaches, assessment instruments must also evolve to reflect current practices and concepts.

Fifth, research examining the relationship between conceptual understanding as measured by the PSCI and practical skills in systematic analysis would provide valuable information about the connection between declarative knowledge and procedural competence. Studies might examine whether students who score higher on the PSCI also perform better on authentic systematic tasks such as constructing phylogenetic trees from morphological or molecular data.

Sixth, investigation of how specific course features relate to misconception reduction could inform course design. Factors such as the use of inquiry-based laboratories, integration of authentic research experiences, emphasis on nature of science concepts, and instructor training in addressing misconceptions might all influence learning outcomes as measured by the PSCI.

As plant systematics continues to evolve through advances in molecular methods and bioinformatics, the need for accurate conceptual understanding becomes ever more critical. Students who lack foundational understanding of classification principles will struggle to engage meaningfully with modern systematic research. The PSCI provides a means of ensuring that students develop the conceptual foundations necessary for advanced study and professional practice in plant sciences.

We encourage educators to use the PSCI as a diagnostic tool at the beginning of instruction to identify prevalent misconceptions, as a formative assessment to monitor learning progress, and as a summative measure

to evaluate instructional effectiveness. The documented validity evidence supports score interpretation for these purposes. We also invite researchers to conduct additional validation studies, examine the instrument's sensitivity to instruction, and contribute to ongoing refinement of this assessment tool. The instrument is available from the corresponding author upon request for educational and research purposes.

References

[1] D. E. Soltis, P. S. Soltis, P. K. Endress, and M. W. Chase, "Phylogeny and Evolution of the Angiosperms," University of Chicago Press, 2018.
[2] J. B. Judd, C. S. Campbell, E. A. Kellogg, P. F. Stevens, and M. J. Donoghue, "Plant Systematics: A Phylogenetic Approach," 4th ed., Sinauer Associates, 2016.
[3] K. Halverson, "Using Pipe Cleaners to Bring the Tree of Life to Life," *American Biology Teacher*, vol. 73, no. 4, pp. 223–227, 2011.
[4] T. R. Gregory, "Understanding Evolutionary Trees," *Evolution: Education and Outreach*, vol. 1, no. 2, pp. 121–137, 2008.
[5] S. A. Baum, S. D. Smith, and S. S. Donovan, "The Tree-Thinking Challenge," *Science*, vol. 310, no. 5750, pp. 979–980, 2005.
[6] R. H. Nehm and L. Reilly, "Biology Majors' Knowledge and Misconceptions of Natural Selection," *BioScience*, vol. 57, no. 3, pp. 263–272, 2007.
[7] E. Meir, J. Perry, J. C. Herron, and J. Kingsolver, "College Students' Misconceptions About Evolutionary Trees," *American Biology Teacher*, vol. 69, no. 7, pp. e71–e76, 2007.
[8] K. Garvin-Doxas and M. Klymkowsky, "Understanding Randomness and Its Impact on Student Learning," *CBE—Life Sciences Education*, vol. 7, no. 2, pp. 227–233, 2008.
[9] M. K. Smith, W. B. Wood, and J. K. Knight, "The Genetics Concept Assessment," *CBE—Life Sciences Education*, vol. 7, no. 4, pp. 422–430, 2008.
[10] D. Hestenes, M. Wells, and G. Swackhamer, "Force Concept Inventory," *The Physics Teacher*, vol. 30, no. 3, pp. 141–158, 1992.
[11] M. K. Smith and J. K. Knight, "Using the Genetics Concept Assessment to Document Persistent Conceptual Difficulties," *Genetics*, vol. 191, no. 1, pp. 21–32, 2012.
[12] K. D. Catley and L. R. Novick, "Seeing the Wood for the Trees," *BioScience*, vol. 58, no. 10, pp. 976–987, 2008.
[13] National Research Council, "Knowing What Students Know," National Academies Press, 2001.
[14] J. Adams and T. G. Wieman, "Development and Validation of Instruments to Measure Learning," *International Journal of Science Education*, vol. 33, no. 9, pp. 1289–1312, 2011.
[15] S. E. Embretson and S. P. Reise, "Item Response Theory for Psychologists," Lawrence Erlbaum Associates, 2000.
[16] C. Henderson, "Common Concerns About the Force Concept Inventory," *The Physics Teacher*, vol. 40, no. 9, pp. 542–547, 2002.
[17] G. J. Posner, K. A. Strike, P. W. Hewson, and W. A. Gertzog, "Accommodation of a Scientific Conception," *Science Education*, vol. 66, no. 2, pp. 211–227, 1982.
[18] R. Driver, "The Pupil as Scientist?" Open University Press, 1983.
[19] P. R. Pintrich, R. W. Marx, and R. A. Boyle, "Beyond Cold Conceptual Change," *Review of Educational Research*, vol. 63, no. 2, pp. 167–199, 1993.
[20] S. Vosniadou, "Capturing and Modeling the Process of Conceptual Change," *Learning and Instruction*, vol. 4, no. 1, pp. 45–69, 1994.
[21] R. J. Mislevy and G. D. Haertel, "Implications of Evidence-Centered Design," *Educational Measurement: Issues and Practice*, vol. 25, no. 4, pp. 6–20, 2006.
[22] F. B. Baker, "The Basics of Item Response Theory," 2nd ed., ERIC Clearinghouse, 2001.
[23] C. C. Hsu and B. A. Sandford, "The Delphi Technique," *Practical Assessment, Research & Evaluation*, vol. 12, no. 10, pp. 1–8, 2007.
[24] P. M. Sadler, "Psychometric Models of Student Conceptions in Science," *Journal of Research in Science Teaching*, vol. 35, no. 3, pp. 265–296, 1998.
[25] D. F. Polit and C. T. Beck, "The Content Validity Index," *Research in Nursing & Health*, vol. 29, no. 5, pp. 489–497, 2006.
[26] J. C. Nunnally and I. H. Bernstein, "Psychometric Theory," 3rd ed., McGraw-Hill, 1994.
[27] K. Halverson, C. Pires, and S. K. Abell, "Exploring the Complexity of Tree Thinking Expertise," *Science Education*, vol. 95, no. 2, pp. 290–313, 2011.
[28] E. Meir, J. Perry, D. Stal, S. Maruca, and E. Klopfer, "How Effective Are Simulated Experiments," *Cell Biology Education*, vol. 4, no. 3, pp. 235–248, 2005.
[29] K. D. Catley, "Darwin's Missing Link," *Science Education*, vol. 90, no. 5, pp. 767–783, 2006.
[30] D. A. Baum and S. D. Smith, "Tree Thinking: An Introduction to Phylogenetic Biology," Roberts and Company Publishers, 2013.
[31] M. W. Klymkowsky, K. Garvin-Doxas, and M. Zeilik, "Bioliteracy and Teaching Efficacy," *Cell Biology Education*, vol. 2, no. 3, pp. 155–161, 2003.
[32] R. Anderson, J. Fisher, and G. Norman, "Development and Evaluation of the Conceptual Inventory of Natural Selection," *Journal of Research in Science Teaching*, vol. 39, no. 10, pp. 952–978, 2002.
[33] K. Halverson, S. K. Abell, M. Friedrichsen, and P. Lederman, "Development and Validation of the Tree-Thinking Concept Inventory," *International Journal of Science Education*, vol. 36, no. 7, pp. 1054–1075, 2014.