

Table S1. Alignment coding of included studies.

Title	Author and Year	Country	P1 Extend ordinary teaching	P2 Participatio n first	P3 Non categorical	P4 Professional judgement	Alignment category
A Face to Face Professional Development Model to Enhance Teaching of Online Research Strategies	Terrazas Arellanes, F. E., Knox, C., Strycker, L.A., & Walden, E. (2016)	United States	1	1	0	1	Partial alignment
Challenges to preparing teachers to instruct all students in inclusive classrooms	Larios, R. J., & Zetlin, A. (2023)	United States	1	1	1	2	Partial alignment
Differentiated instruction in primary mathematics: Effects of teacher professional development on student achievement	Prast, E. J., Van de Weijer Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2018)	Netherlands	0	0	1	1	Limited alignment
Effects of coaching on Universal Design for Learning implementation	Craig, S. L., Smith, S. J., & Frey, B. B. (2022)	United States	2	2	1	2	Stronger alignment
Enhancing Teacher Competence in Differentiated Instruction for English Language Learners with Disabilities: A Professional Development Intervention	Alshaikhi, T., & Khasawneh, M. A. S. (2025)	Saudi Arabia	0	1	0	1	Limited alignment
Enhancing Teachers' Competency through School as Learning Communities and Psychological Capital to Foster Cultural Diversity in Students	Narulmon Prayai, Ruj Luecha, Wissarut Suwansantisuk, Chanita Daengudom (2025)	Thailand	1	1	2	2	Stronger alignment
Exploring Sustained Impacts of Double Check Coaching on Teacher and Student Outcomes: Findings from a Randomized Controlled Trial	Bradshaw, C. P., Ma, J., Pas, E. T., DeCoster, J., Kaihoi, C. A., & Debnam, K. (2025)	United States	1	1	1	1	Stronger alignment
Improving teaching self-efficacy for teachers in inclusive classrooms in Hong Kong	Chao, C. N.G., Forlin, C., & Ho, F. C. (2016)	Hong Kong (China)	0	0	1	0	Limited alignment
Inclusive Education Virtual Professional Development: School Based professionals' Knowledge of Best Practices	Montalbano, C., Lang, J., Coviello, J. C., McQueston, J.A., Hogan, J. A., Nissley Tsiopinis, J., Ciotoli, F., & Buglione, F. (2024)	United States	1	1	1	0	Partial alignment

Nurturing self efficacy for culturally responsive teaching through online language immersion	Courtney, M.B. (2023)	United States	1	1	1	2	Partial alignment
Preliminary Effectiveness of Professional Learning about Disability Specific Evidence Based Classroom Practices for Education Support Staff	Devenish,B.D., Mantilla,A., Bussey, K., McGillivray,J., & Rinehart, N.J. (2023)	Australia	1	0	0	1	Limited alignment
Reaching Out to Paraprofessionals: Engaging Professional Development Aligned With Universal Design for Learning Framework in Rural Communities	Barrio, B.L., & Hollingshead,A. (2017)	United States	1	1	2	2	Partial alignment
School improvement and teachers ' collaborative professional development for inclusive education: A Swedish case	Lelinge, B., & Alwall, J.(2022)	Sweden	1	1	1	2	Partial alignment
Supporting the transition to inclusive education: teachers ' attitudes to inclusion in the Seychelles	Main, S., Chambers, D.J., & Sarah, P. (2016)	Seychell es (Australi a collabora tion)	0	0	1	1	Limited alignment
Improving Science Teaching for All Students	Teresa Arambula Greenfield and Arnold Feldman (1997)	United States	2	2	1	2	Stronger alignment
The Training Program for Individualized Education Programs (IEPs): Its Effect on How Inclusive Education Teachers Perceive their Competencies in Devising IEPs	ilik, \$. \$. , & sari, H. (2017)	Türkiye	0	0	0	1	Limited alignment

Note. Numerical values (0 2) represent qualitative coding of alignment for each principle and are provided for descriptive purposes only; they were not used as a formal scoring system.

Database	Date searched	Search string	Limits / filters applied
ERIC	[2026-02-05]	("inclusive pedagogy" OR "inclusive teaching" OR "inclusive classroom practice" OR "teaching for all" OR "inclusive pedagogical skills" OR "universal design for learning" OR "culturally responsive teaching" OR "differentiated	Peer-reviewed journal articles; English

		instruction") AND ("teacher training" OR "professional development" OR "professional learning" OR "in-service teacher*" OR "teacher workshop*")	language
Web of Science Core Collection	[2026-02-05]	TS=("inclusive pedagogy" OR "inclusive teaching" OR "inclusive classroom practice" OR "teaching for all" OR "inclusive pedagogical skills" OR "universal design for learning" OR "culturally responsive teaching" OR "differentiated instruction") AND TS=("teacher training" OR "professional development" OR "professional learning" OR "in-service teacher*" OR "teacher workshop*")	Document type: Article; source type: Journal; language: English
Scopus	[2026-02-05]	TITLE-ABS-KEY(("inclusive pedagogy" OR "inclusive teaching" OR "inclusive classroom practice" OR "teaching for all" OR "inclusive pedagogical skills" OR "universal design for learning" OR "culturally responsive teaching" OR "differentiated instruction") AND ("teacher training" OR "professional development" OR "professional learning" OR "in-service teacher*" OR "teacher workshop*"))	Document type: Article; source type: Journal; language: English

Table S2. Full Search Strategies

Table S3. Detailed Outcome Data and Reported Effect Indicators Across Included Studies

Study	Country	Duration	Design / comparison	Teacher-level outcomes	Student-level outcomes	Classroom / school-level outcomes	Reported effect indicators
Terrazas-Arellanes et al. (2016)	United States	2 academic years	Single-group quasi-experimental	Practice: 100% of teachers reported growth in differentiated instruction and technology integration.	Online research skills improved overall: 31.3 (22.1) → 60.5 (23.0), ***p* < .001. Students with learning disabilities:	Not reported.	Not reported.

Study	Country	Duration	Design / comparison	Teacher-level outcomes	Student-level outcomes	Classroom / school-level outcomes	Reported effect indicators
					19.0 (18.9) → 49.7 (25.4). General education students: 34.0 (22.0) → 63.0 (21.9).		
Larios & Zetlin (2023)	United States	2 academic years	Single-group mixed methods	Beliefs: teachers' attitudes shifted from "I can't" to "how can we?" with increased shared responsibility. Self-efficacy: 89.5% reported feeling more prepared. Practice: teachers implemented differentiated instruction and classroom management strategies.	Not reported.	PLCs were perceived as effective in improving school collaboration culture.	Not reported.
Prast et al. (2018)	Netherlands	1 academic year	Quasi-experimental cohort design with comparison group	Beliefs: mean ratings for perceived effectiveness of differentiated instruction were above 3.4. Knowledge: professional knowledge scores were above midpoint. Self-efficacy: Cohort 1 M = 3.84; Cohort 2 M = 4.00. Practice: 76–77% reported actual application of differentiated instruction	Mathematics achievement: Year 1 Cohort 1 vs. control $\beta = 0.15$, $p = .007$; Year 2 Cohort 2 vs. control $\beta = 0.03$, $p = .640$; Year 2	Not reported.	Student achievement effect was significant only in Year 1.

Study	Country	Duration	Design / comparison	Teacher-level outcomes	Student-level outcomes	Classroom / school-level outcomes	Reported effect indicators
				strategies.	follow-up Cohort 1 $\beta = 0.06$, reported $p = -.665$.		
Craig et al. (2022)	United States	9–14 weeks	Quasi-experimental with comparison group	Beliefs: coaching was perceived as non-judgemental and collegial. Knowledge: UDL knowledge effect was non-significant, $F(1,37) = 0.231$, $p = .633$. Self-efficacy: increased confidence for implementing UDL. Practice: self-reported UDL implementation increased, but coaching effects were non-significant, $F(1,37) = 0.010$, $p = .921$.	Not reported.	Observed UDL implementation increased directionally, but the interaction effect was non-significant, $F(1,16) = 3.137$, $p = .096$, $\eta^2 = .164$.	$\eta^2 = .164$ for observed UDL implementation.
Alshaikhi & Khasawneh (2025)	Saudi Arabia	1 academic year	Single-group pre-post	Beliefs: perceived usefulness was not significantly associated with competence gains, $r = -0.04$, $p = .831$. Knowledge/self-efficacy: competence measure included knowledge and self-efficacy, both improved. Practice: differentiated instruction competence improved from	Not reported.	Not reported.	Significant improvement in self-reported differentiated instruction competence.

Study	Country	Duration	Design / comparison	Teacher-level outcomes	Student-level outcomes	Classroom / school-level outcomes	Reported effect indicators
				M = 3.32, SD = 1.14 to M = 3.78, SD = 1.14, paired t = -9.64, p < .001.			
Prayai et al. (2025)	Thailand	1 academic year	Single-group quasi-experimental	Practice: median scores for four teaching competency dimensions reached Level 3 and were significantly higher than the Level 2 criterion; all $z \geq 3.74$, p < .01.	Cultural diversity understanding improved: Wilks' $\Lambda = .28$, F = 148.68, p < .001, $\eta^2 = .72$. Gender diversity: 2.59 → 3.46; sexual orientation: 2.66 → 3.54; race/religion: 2.59 → 3.56; social class: 2.57 → 3.56; special needs: 2.60 → 3.60.	Not reported.	$\eta^2 = .72$ for multivariate student outcome.
Bradshaw et al. (2025)	United States	1 year + 1-year follow-up	Randomised controlled trial	Teacher self-report: culturally responsive teaching self-efficacy improved from pre-test to post-test, but the effect	CLASS-S student engagement improved from pre-	CLASS-S classroom interaction outcomes improved	CLASS-S emotional support and instructional support showed sustained effects; ASSIST indicators were non-significant.

Study	Country	Duration	Design / comparison	Teacher-level outcomes	Student-level outcomes	Classroom / school-level outcomes	Reported effect indicators
				became marginal at follow-up. No sustained effects were found for multicultural efficacy, classroom behaviour management efficacy, or teacher stress.	test to post-test, but the effect was not sustained at follow-up. ASSIST student behaviour indicators were not significant.	from pre-test to post-test across emotional support, instructional support, classroom organisation, and student engagement. Emotional support and instructional support were sustained at follow-up. ASSIST observed behavioural indicators were not significant.	
Chao et al. (2016)	Hong Kong, China	30 hours	Single-group mixed methods	Knowledge: policy knowledge improved from 1.36 (0.58) to 2.10 (0.70), $d = 1.15$. Self-efficacy: teaching confidence improved from 2.68 (0.67) to 3.02 (0.55), $d = 0.55$; behaviour management $d = 0.37$; collaboration $d =$	Not reported.	Not reported.	$d = 1.15, 0.55, 0.37, 0.52,$ and 0.28 .

Study	Country	Duration	Design / comparison	Teacher-level outcomes	Student-level outcomes	Classroom / school-level outcomes	Reported effect indicators
				0.52; inclusive teaching $d = 0.28$.			
Montalban o et al. (2024)	United States	4–6 weeks	Single-group repeated measures	Knowledge: co-teaching improved from 0.68 to 0.74, Est. = 0.06, $p < .001$; differentiated instruction improved from 0.70 to 0.81, Est. = 0.11, $p < .001$; UDL improved from 0.68 to 0.89, $t = 9.24$, $p < .001$, $d = 1.57$; accommodations/modifications improved from 0.57 to 0.75, $t = 8.68$, $p < .001$, $d = 1.16$.	Not reported.	Not reported.	$d = 1.57$ for UDL; $d = 1.16$ for accommodations/modifications.
Courtney (2023)	United States	80 hours	Single-group pre-post	Beliefs/self-efficacy: culturally responsive teaching outcome expectancy increased from $M = 86.1$ to $M = 91.3$, but was non-significant, $t = -1.2$, $p > .05$, $d = 0.32$. Culturally responsive teaching self-efficacy improved from $M = 64.1$ to $M = 82.3$, $t = -3.0$, $p < .01$, $d = 0.86$. Culturally responsive classroom management self-efficacy improved from $M = 72.4$ to $M = 84.0$, $t = -1.8$, $p < .05$, $d = 0.59$.	Not reported.	Not reported.	$d = 0.86$ for CRTSE; $d = 0.59$ for CRCMSE; $d = 0.32$ for CRTOE, non-significant.

Study	Country	Duration	Design / comparison	Teacher-level outcomes	Student-level outcomes	Classroom / school-level outcomes	Reported effect indicators
Devenish et al. (2023)	Australia	4–6 hours	Single-group repeated measures	Self-efficacy: primary teachers increased from median 3.00 to 3.50, $Z = -9.19$, $p < .001$, $r = 0.62$; secondary teachers increased from 3.00 to 3.50, $Z = -5.77$, $p < .001$, $r = 0.62$. Knowledge: primary support staff increased from median 3.80 to 4.20, $Z = -10.12$, $p < .001$, $r = 0.68$; secondary staff increased from 4.00 to 4.67, $Z = -6.72$, $p < .001$, $r = 0.71$. Practice: 19% expressed intention to apply learned strategies; no observed behavioural data.	Not reported.	Not reported.	$r = 0.62$ – 0.71 .
Barrio & Hollingshead (2017)	United States	1 academic year	Single-group mixed methods	Beliefs: positive attitudes toward inclusion and collaboration were reported. Knowledge: qualitative themes indicated gains in student assessment, service provision, software/resources, and curriculum standards. Practice/self-efficacy: high post-test ratings reflected perceived effectiveness and perceived growth.	Not reported.	Not reported.	Not reported.

Study	Country	Duration	Design / comparison	Teacher-level outcomes	Student-level outcomes	Classroom / school-level outcomes	Reported effect indicators
Lelinge & Alwall (2022)	Sweden	2 years	Single-group mixed methods	Beliefs: school climate and satisfaction improved, including participation in decision-making 7% → 41%, shared responsibility 4% → 22%, collaboration culture 11% → 46%, liking school 36% → 60%, recommendation 22% → 57%, job satisfaction 31% → 43%. Self-efficacy: ability to manage disruptive behaviour increased from 38% to 53%, but several confidence indicators declined. Knowledge: PD needs for behaviour management, personalised learning, and special needs teaching decreased. Practice: team teaching increased from 62% to 66%, but classroom observation declined from 29% to 16%.	Not reported.	Collaboration culture and organisational change improved; feedback culture also improved.	Not reported.
Main et al. (2016)	Seychelles	1 semester	Single-group pre-post	Beliefs: SACIE-R total score $d = 0.75$; attitudes $r = 0.54$. Self-efficacy: TEIP total score $d = 0.38$; collaboration efficacy $r = 0.92$; behaviour	Not reported.	Not reported.	$d = 0.38$ and 0.75 ; $r = 0.92$, 0.54 , 0.88 .

Study	Country	Duration	Design / comparison	Teacher-level outcomes	Student-level outcomes	Classroom / school-level outcomes	Reported effect indicators
				management declined significantly, $r = 0.88$. Practice: designing learning tasks improved from 3.75 to 5.06; collaboration declined from 5.06 to 4.45.			
Arámbula-Greenfield & Feldman (1997)	United States	1 semester	Quasi-experimental with comparison group	Self-efficacy: confidence in addressing gender inequity increased from 3.083 to 3.833, $t = -2.462$. Knowledge: content knowledge increased from 3.250 to 3.750, $t = -1.593$, non-significant. Practice: observations indicated a shift from lecture-based instruction to hands-on experimental teaching, with greater use of gender-neutral language and structured cooperative learning.	Gender stereotypes, perceived ability, and achievement showed no significant between-group differences. On-task behaviour increased from 59.4% to 77.5%, $\chi^2 = 37.22$, $p \leq .05$; girls: 57.5% $\rightarrow$ 83.4%, $\chi^2 = 33.72$; boys: 61.2% $\rightarrow$ 67.7%, $\chi^2 = 7.94$.	Not reported.	Significant observed gains in on-task behaviour; content knowledge non-significant.
İlik & Sarı (2017)	Türkiye	32 hours	Experimental/control comparison	Beliefs: 79% believed evaluation was necessary; all participants agreed	Not reported.	Not reported.	Significant post-test group difference in perceived IEP competence.

Study	Country	Duration	Design / comparison	Teacher-level outcomes	Student-level outcomes	Classroom / school-level outcomes	Reported effect indicators
				support services should be provided and barriers to collaboration removed. Self-efficacy: post-test IEP competence was significantly higher in the experimental group than in the control group, $M = 4.60$ vs. 3.09, $t = 9.1$, $p < .05$. Knowledge: qualitative findings showed persistent information-access and team-functioning challenges. Practice: IEP evaluation practices were often superficial; only a small number of teachers could evaluate students according to individual levels.			

Note. Bel = beliefs; Know = knowledge; SE = self-efficacy; Pract = instructional practice; UDL = Universal Design for Learning; IEP = Individualized Education Program; PLC = professional learning community; CRTSE = Culturally Responsive Teaching Self-Efficacy; CRCMSE = Culturally Responsive Classroom Management Self-Efficacy; CRTOE = Culturally Responsive Teaching Outcome Expectancy; CLASS-S = Classroom Assessment Scoring System–Secondary; ASSIST = Assessing School Settings: Interactions of Students and Teachers. Where possible, results are reported in the form provided by the original studies. “Not reported” indicates that the outcome domain was not reported in the reviewed article.

Table S4. Detailed Outcome Data and Effect Indicators Across Included Studies

Author and Year	Country	Duration	Comparison	Design	Teacher-level	Student-level	Classroom/School-level	Effect size
Terrazas-Arellanes (2016)	United States	2 academic years	Single-group	Quasi-exp	Pract: 100% reported growth in differentiated instruction and technology integration	Online research skills: Overall 31.3 (22.1) → 60.5 (23.0)***; SWLD 19.0 (18.9) → 49.7 (25.4); general education students 34.0 (22.0) → 63.0 (21.9)	/	/
Larios & Zetlin (2023)	United States	2 academic years	Single-group	Mixed-methods	Bel: Cross-disciplinary collaboration improved morale; teachers' attitudes shifted from "I can't" to "how can we?", with increased shared responsibility and reduced feelings of isolation and helplessness. SE: 89.5% of teachers reported feeling "more prepared" to meet students' needs. Pract: Teachers implemented multiple new differentiated instruction strategies, including small-group instruction to support students with ADHD, and classroom management strategies such as token boards, visual schedules, and timers to increase engagement and reduce behavioural problems. UDL became a topic of discussion in meetings, and teachers recognised that they had already been using some UDL principles implicitly, with workshops helping to clarify these practices.	/	School collaboration culture—PLCs perceived as effective	/

Prast et al. (2018)	Netherlands	1 academic year	Comparison-group	Quasi-exp (cohort)	Bel: Mean ratings for beliefs about the effectiveness of differentiated instruction were all above 3.4, indicating positive expectations. Know: Mean scores for differentiated instruction professional knowledge were all above 3 (midpoint = 3). SE: Self-efficacy for implementing differentiated instruction: Cohort 1 M = 3.84, Cohort 2 M = 4.00. Pract: 6–77% of teachers reported “actual application” of differentiated instruction strategies.	Mathematics achievement: Year 1 Cohort 1 vs control $\beta = 0.15$, $p = .007$; Year 2 Cohort 2 vs control $\beta = 0.03$, $p = .640$; Year 2 follow-up Cohort 1 $\beta = -0.06$, $p = .665$.	/	/
Craig et al. (2022)	United States	9–14 weeks	Comparison-group	Quasi-exp (comparison)	Bel: Teachers emphasised the importance of the coaching relationship; coaching was perceived as non-judgemental, safe professional support and as collegial collaboration rather than evaluation. Coaching experience was considered important, and coaching was recognised as an effective form of PD for understanding and implementing UDL. Know: UDL knowledge, intervention effect non-significant, $F(1,37) = 0.231$, $p = .633$. SE: Increased self-efficacy/confidence for implementing UDL. Pract: Self-reported UDL implementation increased from pre- to post-test, but coaching effects were non-significant, $F(1,37) = 0.010$, $p = .921$. Observational data also showed increased implementation, but the interaction effect was non-significant, $F(1,16) = 3.137$, $p = .096$, $\eta^2 = .164$; supplementary t-tests indicated a near-significant post-test group difference ($p = .15$).	/	/	0.164

Alshaikhi & Khasawneh (2025)	Saudi Arabia	1 academic year	Single-group	Pre-post	Bel: Perceived usefulness of the intervention was not significantly associated with post-test competence, $r = -0.04$, $p = .831$. Know/SE: The competence measure included both knowledge and self-efficacy, both of which improved significantly. Pract: Self-reported differentiated instruction competence improved significantly from pre-test ($M = 3.32$, $SD = 1.14$) to post-test ($M = 3.78$, $SD = 1.14$), paired $t = -9.64$, $p < .001$.	/	/	/
Prayai et al. (2025)	Thailand	1 academic year	Single-group	Quasi-exp	Pract: Median scores for all four dimensions of teaching competency (learning activity design, instructional implementation, instructional media selection, and assessment and evaluation) reached Level 3 and were all significantly higher than the predefined Level 2 criterion (all $z \geq 3.74$, $p < .01$).	Academic achievement/subject-related skills—understanding of cultural diversity: MANOVA indicated a significant multivariate effect, Wilks' $\Lambda = .28$, $F = 148.68$, $p < .001$, $\eta^2 = .72$. All dimensions improved significantly ($p < .001$): gender diversity $2.59 \rightarrow 3.46$; sexual orientation $2.66 \rightarrow 3.54$; race and religion $2.59 \rightarrow 3.56$; social class $2.57 \rightarrow 3.56$; individuals with special needs $2.60 \rightarrow 3.60$.	/	/
Bradshaw et al. (2025)	United States	1 year + 1-year follow-	Comparison-group	RCT	SE: Significant intervention-group gains in culturally responsive teaching self-efficacy from pre-test to post-test (p-adjusted significant); post-	Student engagement: Intervention-group gains from pre-test to post-test	/	0.86 (CRTSE) 0.59

		up			test to follow-up and pre-test to follow-up effects became marginally significant ($p = .064$). No significant effects were found for multicultural efficacy, classroom behaviour management efficacy, or work stress. Coaches' MI-consistent language was positively associated with teachers' culturally responsive teaching self-efficacy and multicultural efficacy, and negatively associated with stress (significant across all three time periods). Teacher change talk was negatively associated with stress (significant). Coach-reported barriers were negatively associated with efficacy gains (significant). Pract (CLASS observations): From pre-test to post-test, all four dimensions—emotional support, instructional support, classroom organisation, and student engagement—showed significant positive immediate effects (p-adjusted significant). No significant effects were found from post-test to follow-up. From pre-test to follow-up, emotional and instructional support remained significant, whereas classroom organisation and student engagement did not. Pract (ASSIST observations): No significant effects were found across all 10 dimensions from pre-test to post-test. From post-test to follow-up and pre-test to follow-up, counterintuitive findings emerged for disruptive behaviour, with larger declines in the control group than in the intervention group.	were significantly greater than those in the control group (p-adjusted significant); no sustained effects were found from post-test to follow-up or from pre-test to follow-up.		(CRCMSE)
Chao et al. (2016)	Hong Kong (China)	30 hours	Single-group	Mixed-methods	Know: Policy knowledge 1.36 (0.58) → 2.10 (0.70), $d = 1.15$. SE: Teaching confidence 2.68 (0.67) → 3.02 (0.55), $d = 0.55$; behaviour	/	/	1.15 (policy knowledge) 0.55

					management $d = 0.37$; collaboration $d = 0.52$; inclusive teaching $d = 0.28$.			(teaching confidence) 0.37 (behaviour management) 0.52 (collaboration) 0.28 (inclusive teaching)
Montalbano et al. (2024)	United States	4–6 weeks	Single-group	Repeated measures	Know: Co-teaching ($n = 102$, 6 schools) improved from pre-test 0.68 to post-test 0.74, with a significant time effect (Est. = 0.06, $p < .001$) and non-significant school effects; differentiated instruction ($n = 211$, 13 schools) improved from 0.70 to 0.81, with a significant time effect (Est. = 0.11, $p < .001$) and non-significant school effects; UDL ($n = 54$) improved from 0.68 to 0.89, $t = 9.24$, $p < .001$, $d = 1.57$; accommodations and modifications ($n = 64$) improved from 0.57 to 0.75, $t = 8.68$, $p < .001$, $d = 1.16$.	/	/	1.57 (UDL) 1.16 (Acc_Mods)
Courtney (2023)	United States	80 hours	Single-group	Pre-post	Bel/SE: Culturally responsive teaching outcome expectancy (CRTOE) increased from pre-test $M = 86.1$ to post-test $M = 91.3$, $t = -1.2$, $p > .05$, $d = 0.32$, non-significant; median pre-test scores were already above 90, suggesting a possible ceiling effect. SE: Culturally responsive teaching self-efficacy (CRTSE) increased from pre-test $M = 64.1$ to post-test $M = 82.3$, $t = -3.0$, $p < .01$, $d = 0.86$; culturally responsive classroom management self-efficacy (CRCMSE) increased from pre-test $M = 72.4$ to post-test $M = 84.0$, $t = -1.8$, $p < .05$, $d = 0.59$.	/	/	0.86 (CRTSE) 0.59 (CRCMSE) 0.32 (CRTOE, n.s.)

Devenish et al. (2023)	Australia	4–6 hours	Single-group	Repeated measures	SE/Bel: Qualitative feedback reflected improvements in confidence and beliefs. SE: Primary teachers (n = 221) showed median scores increasing from 3.00 to 3.50, $Z = -9.19$, $p < .001$, $r = 0.62$; secondary teachers (n = 87) increased from 3.00 to 3.50, $Z = -5.77$, $p < .001$, $r = 0.62$. Know: Primary support staff (n = 222) increased from median 3.80 to 4.20, $Z = -10.12$, $p < .001$, $r = 0.68$; secondary staff (n = 89) increased from 4.00 to 4.67, $Z = -6.72$, $p < .001$, $r = 0.71$. Pract: 19% of participants expressed intentions to apply the learned strategies in future practice (intention-level only; no observed behavioural data).	/	/	/
Barrio & Hollingshead (2017)	United States	1 academic year	Single-group	Mixed-methods	Bel: Positive attitudes toward inclusion and collaboration were reported. Pract/SE: High post-test ratings reflected perceived effectiveness, indirectly supporting improvements in instructional practice and self-efficacy. Know: Qualitative themes indicated gains in knowledge and skills related to student assessment, service provision, educational software/resources, and curriculum standards. Pract: Pre-test interest ratings (5-point scale)—integrating technology/strategies into the classroom $M = 3.11$–3.67; knowledge of available resources $M = 3.22$–4.00; practical guidelines $M = 2.92$–3.90; procedural/technology knowledge $M = 2.72$–3.38. Post-test perceived gains—classroom usefulness $M = 4.10$–4.61; integration of critical thinking and technology $M = 3.10$–4.44; deeper consideration of student learning outcomes $M =$	/	/	/

					4.10–4.67; effective use of time M = 4.15–4.72.			
Lelinge and Al-wall (2022)	Sweden	2 years	Single-group	Mixed-methods	Bel: Substantial improvements in school climate and satisfaction: participation in decision-making 7% → 41%, shared responsibility 4% → 22%, collaboration culture 11% → 46%, additional support 11% → 32%, liking school 36% → 60%, school recommendation 22% → 57%, and job satisfaction 31% → 43%. Evaluation and feedback also improved: perceived limited impact of evaluation declined from 22% to 8%, comprehensive evaluation feedback increased from 20% to 55%, and discussion of correcting weaknesses increased from 49% to 61%. SE: Ability to manage disruptive classroom behaviour increased from 38% to 53%; however, several other teaching confidence indicators declined, including calming noisy students (51% → 42%), motivating low-interest students (38% → 26%), using multiple forms of assessment (29% → 26%), providing alternative explanations (58% → 42%), using alternative strategies (44% → 37%), and satisfaction with school performance (31% → 27%). Know: PD needs related to behaviour management, personalised learning, and special needs teaching decreased (16% → 3%, 29% → 3%, and 33% → 11%, respectively). Positive effects of completed PD were reported for behaviour management (80% → 94%), personalised learning (56% → 63%), and special needs teaching (70% → 63%). Pract: Team teaching increased from 62% to 66%,	/	Collaboration culture and organisational change: substantial improvements in school climate (e.g., participation in decision-making 7% → 41%, collaboration culture 11% → 46%); feedback culture also improved, with the proportion reporting little impact of feedback declining from 22% to 8%.	/

					whereas classroom observation of colleagues declined from 29% to 16%, and student development discussions declined from 62% to 40%. Feedback focus increased for behaviour management (42% → 61%), providing improvement-oriented feedback (22% → 50%), and collaboration (40% → 68%). Reported effects of feedback increased for job satisfaction (24% → 58%) and motivation (31% → 58%), whereas perceived effects on special needs teaching methods declined (22% → 13%).			
Main et al. (2016)	Seychelles	1 semester	Single-group	Pre-post	Bel: SACIE-R total score $d = 0.75$; attitudes $r = 0.54$. SE: TEIP total score $d = 0.38$; collaboration efficacy $r = 0.92$; behaviour management declined significantly $r = 0.88$. Pract: Designing learning tasks improved from 3.75 to 5.06; collaboration declined from 5.06 to 4.45.	/	/	0.38 (TEIP total score) 0.75 (SACIE-R total score) 0.92 (collaboration efficacy) 0.54 (attitudes)
Arámbula-Greenfield and Feldman (1997)	United States	1 semester	Comparison-group	Quasi-exp	SE: Confidence in addressing gender inequity strategies increased significantly from 3.083 to 3.833, $t = -2.462$; no significant changes were found in the control group. Know: Content knowledge related to gender inequity increased from 3.250 to 3.750, $t = -1.593$, non-significant; no significant changes were found in the control group. Pract: Qualitative observations indicated that in-service teachers shifted from lecture-based instruction to hands-on experimental teaching, with greater use of gender-neutral language and	Gender stereotypes: Comparisons of change scores between the intervention and control groups showed no significant between-group differences in gender stereotypes, perceived ability, or achievement. Classroom observations: Overall on-task behaviour increased from 59.4% to 77.5% ($\chi^2 = 37.22$, $p \leq$	/	/

					structured cooperative learning groups.	.05); girls 57.5% → 83.4% ($\chi^2 = 33.72$); boys 61.2% → 67.7% ($\chi^2 = 7.94$). Off-task behaviour declined more substantially among girls, particularly in classrooms of novice teachers.		
İlik and Sarı (2017)	Türkiye	32 hours	Comparison-group	Experimental (control)	Bel: 79% believed evaluation was necessary to support subsequent teachers, and all participants agreed that support services should be provided and negative factors affecting collaboration should be eliminated. SE: Perceived IEP competence showed no pre-test differences between groups ($M = 3.13$ vs 3.12), whereas post-test scores were significantly higher in the experimental group than in the control group ($M = 4.60$ vs 3.09, $t = 9.1$, $p < .05$). Know: Qualitative findings showed that nearly half of teachers experienced difficulties accessing information; approximately 46% reported that teams existed only on paper; 75% had not formally defined performance domains; 67% set goals without consulting others; 83% reported no difficulty defining services but experienced barriers in accessing them; and 54% believed that parents required education. Pract: Qualitative evaluation of IEP practices showed that 63% considered semester and end-of-year evaluations superficial; 79% viewed evaluation as necessary; 63% reported that	/	/	/

					inappropriate goals prevented evaluation results from being included in reports; and only a small number of teachers were able to evaluate students according to their individual levels.			
--	--	--	--	--	---	--	--	--