Project Title	**:**	Adapting task-based pedagogy for students with special educational needs: Training, practising and validation
Grantee	**:**	Hong Kong Baptist University
Principal Investigator	**:**	CHAN Hang, Joshua Hong Kong Baptist University
Co-Investigator(s)	**:**	CHOI Tat Heung Hong Kong Baptist University

Final Report

by

Principal Investigator

(a) **Title**

Adapting task-based pedagogy for students with special educational needs: Training, practising and validation

(b) Abstract

Task-based language teaching (TBLT) assumes that general learners can benefit from doing tasks by negotiating meaning, making sense of them, experimenting with words and sentences that eventually become part of their proficiency and communicative competence (Mackey, 2012; Pica, 1996;). Therefore, tasks will become more challenging as they are sequentially sequenced more strategically and cognitive reasoning skills are gradually developed (Robinson, 2015). Over the past 30 years, these viewpoints have not changed much, but this implicit assumption presents fundamental challenges in the context of special education (SE) classrooms, in which there are students with dyslexia, autism, or cognitive impairment, who have difficulty reasoning, using literacy skills, or communicating verbally (Daloiso, 2017; Hockly, 2016). This research presents a study between 2021-2024 that attempted to adapt TBLT to SEN classrooms. This study evaluated the implementation of TBLT by seven teachers in six SEN schools, analysing 91 lessons with 273 times of observations - possibly the first study to assess this group's English language learning in Hong Kong. Among the outcomes, a 5-dimensional TBLT framework for lesson planning and evaluation has been proposed. The framework advocate moving away from textbook tasks or negotiating external language forms, but promote a direct experience to concept development and language learning, via (a) a context in which experience, emotions and feelings are encouraged, (b) well-defined input, © encouraging output via (d) teachers' improvisational techniques, and (e) utilising assistive techniques. Based on multiple regressions, the 5-dimensional framework explains 45.8% of the variance in task-basedness in the teachers' lessons and explains 61.5% of the variance. Partial Least Squares Structural Equation Modeling (PLS-SEM) was used to condense the instrument, keeping the salient questions and eliminating those with poor loadings. This report offers suggestions for why the framework might not be fully utilised and proposes reunderstanding of known definitions of TBLT.

(c) Keywords

Task-based Language Teaching (TBLT), special schools, learners with intellectual disabilities, 5-dimensional TBLT framework

(d) Introduction

The study's prequel: In the application document, the project team proposed a 5-dimensional class observation scheme for Special Educational Needs (SEN) learners, focusing on ecological-semiotic sphere (larger environment), task design, task meaning, tools and mediation, and assistive techniques. This concept is different from existing TBLT paradigms that emphasise micromanagement of task factors (e.g., accuracy, fluency, and complexity) and manipulation of forms, but instead emphasise a 'deep-dive' context that promotes personal investment, purpose and thinking (Bolton, 1977; Lambert, Aubrey, & Bui, 2023), clear input and teacher's scaffolding. Building on this wish, this study aims to test this framework (left) in order to develop a practical schema for teachers' lesson planning in SEN classrooms.

The primary reason of proposing a deep dive into language, social and emotional learning (LSEL) is to fill the gaps left by other models, such as Ellis's (2005) principle-based framework, Willis and Willis's (1996) staged framework (pre-task), Seedhouse's (2005) workplan-or-process framework, and Skehan's (1998) cognitive framework (task as a means to allocate resources), Long's (2015) task as a syllabus model (2015). These models describe and dissect the components of a task, or measure changes in task-doing processes. Apart from Willis and Willis, which presents ideas about task stages, few have devoted time to this aspect, or their focus has been on narrow aspects. The importance of tasks for personal investment, language development, and emotional development is not emphasised until recently by Lambert (2023). Developing a framework that emphasises sequential planning and engagement considerations for promoting LSEL is therefore crucial.

Table 1. Framework in the original proposal

Proposed task-design framework in the proposal	Observation scheme in the proposal

Table 3. The five-dimensional TBL framework

Dimensions	Examples*	Theories
Dimension 1 (Ecological-semiotic sphere) **D1.1 Story as ecology** A story is a macro-organiser of an ecology. Every learner is an active agent capable of picking up meanings in pursuit of some learning targets. **D1.2 Three worlds** Lessons should lead	A story as an ecology serves to interrelate a series of tasks. In this story, the students were on a journey to look for a missing mother; ↓  In another story, the teacher asked the students to interact with an English-speaking	• Drawing on ideas from systemic-functional linguistic theory and ecological-semiotic theory, TBL should create a classroom ecology that emphasises the relationship between language and the wider world in which language is used (Halliday, 1978; van Lier, 2004a, 2008). • Tasks are not merely language tasks, but also

THIS SCHEME IS FOR A SINGLE LESSON

	Guiding questions	Pedagogies	Comments
Dimension 1	Is the story coherent and relatable for students with mild ID? Does the story involve the three worlds?	**D1.1 Story as ecology** ☐ *The story has **a clear sociocultural setting** (e.g., utilises the community as a resource).* ☐ *The story is coherent and **inter-relates tasks.*** **D1.2 Three worlds** ☐ *The story introduces students to the **physical world** (in a concrete scenario).* ☐ *Introduces students to the **social world** (i.e., includes people's interactions)* ☐ *Introduces students to the **semiotic world** (enables students to make meanings via language and non-language means).*	
Dimension 2	Does the teacher include a variety of TBL tasks? What is the spatio-temporal structure of each task?	**D2.1 Mission-in-motion** ☐ *A **mission** drives the students forward.* ☐ *The teacher uses that mission to **motivate** the students to complete a task.* **D2.2 TBL tasks** ☐ *There is a variety of **TBL tasks** (e.g., role-play, information-gap, problem-solving, or opinion-exchange tasks).* ☐ *There is a variety of **interactional patterns** (e.g., group work, pair work, team interactions, etc.)* **D2.3 Task internal structure** ☐ Tasks include a ***setting, people,*** and involves a *mode of communication (e.g., reading, writing*	

(e) Review of Literature of the Project

Meta-analyses flourished: Task-Based Language Teaching (TBLT) has emerged as a researched pedagogy that consistently shows effectiveness across multiple learning settings. Several meta-analyses of TBLT have evaluated its efficiency. Boers and Faez (2023) have reported an aggregated effect size of d=0.93, which is very close to Keck et al (2006) d=0.92 after an analysis of 14 TBLT studies on the acquisition of lexical and grammatical items. These studies studied traditional tasks, including jigsaw, information gap, problem-solving, decision-making, opinion exchange, and narrative. Bryfonski and McKay's (2019) meta-analysis that claims to have a broader scope than Keck, Iberri-Shea, Tracy-Ventura, and Wa-Mbaleka (2006) includes 52 studies; the study concludes that TBLT has medium to large effects. Bryfonski's study is disputed by Boers and Faez's (2023) study, who question the former study for including studies that should not have been included (due to the definition of task and the experimental methods). Overall, the positive effects of TBLT have prompted more interest in short exchange tasks (such as jigsaws or problem-solving), because they are controllable, comparable, and compact for measurement, benefiting the researcher's agenda. However, the overemphasis on balance, granularity, and micro-features has done little to encourage practitioners to consider the broader educational benefits of TBLT, such as its learner-centred nature, or discovery tasks, that sadly are only available in edited books or found only in older texts.

Task-based research in SEN contexts: Understanding new parameters

One prevailing attitude in SLA research today is: Teachers are allowed to engage in a wide range of TBLT programmes, but when it comes to reporting research, scientific rigour is paramount. In line with what Sato and Loewen (2022) mentioned, research and teaching need to contribute more reciprocally, especially in the context of special education where "practitioners can inform L2 research to make it more relevant for the classroom." (p.509). To date, niches or advocates of task research may be more interested in how task-structure prompts engagement or attention-directing mechanisms in task designs (Hiver & Wu, 2023; Philp & Duchesne, 2016; Reeve & Lee, 2014), has promoted multidimensional kinds of engagement and attention-directing engagement during TBLT. These interesting ideas are related to our framework. To date, interest in newer niches of task research are only confined to smaller circles. Teachers searching for TBLT literature mainly find short exchange tasks or broad task principles. To give an example: one broad principle that defines a task as 'an activity with a communicative focus and a non-linguistic outcome' does not adequately describe the work of SEN teachers. How is communicative focus when students have difficulty communicating? Suggestions which are either too specific or too broad do not provide ecologically valid insights ready for application in SEN classrooms.

The second problem with introducing TBLT into the SEN context is not that it is impossible, but it requires balancing the TBLT principles with students' limited abilities. This requires re-contextualising new parameters, such as LwSENs' limited attention spans, their inflexible thinking, school's lengthier unit duration with few assessment requirements. This demands a grounded approach to research (Glaser, 1978), where what we theorise will be based on the data we collect (Merriam, 1998). A third problem with current TBLT research relates to its sampling biases: Godfroid and Andringa (2023) in a recent special issue have questioned how SLA have sourced their samples exclusively from university settings which does not reflect the needs of non-academic participants. This sampling bias problem overlooks the needs, for example, of what Tarone (2010) calls low-literate learners. These learners share certain literacy characteristics with learners with SEN. The learners in this study are further down the

continuum taking a separate curriculum in special schools. Overall, recent TBLT meta-analyses indicate that researchers analyse task conditions microscopically and generate insights at a micro-level levels (but with much scientific rigour). Although the micro-research agendas do not always align, they are all in agreement about the importance of learner-centredness, meaning-making, and goal-orientedness in doing tasks.

Concepts to be investigated in this study: Criticism of TBLT research does not imply that it is unimportant for the research of SEN. One well-known concept in TBLT is *negotiation for meaning* (Long, 1996), a special type of interaction that involves clarification, feedback, forced reformation, and strategic repairs. However, this concept (as well as 'meaning-making') does not apply to LwSENs because they do not strategically repair, adapt, and then communicate. To apply it to SEN classrooms, this concept requires clarification, such as if negotiation processes can be applied in their entirety, or in part. Consequently, readjusting our expectations and measurement methods is an important question this study should address. Second, modern TBLT theories champion the idea that task configuration (e.g., planning time, rubrics) can change task requirements, thereby fostering language acquisition or language reformulation. The assumption that tasks or learners can co-adapt themselves cannot be applied straightforwardly to learners with SENs. Finally, the macro-goal of TBLT is to promote an unconscious process of meaningful communication; this goal is in line with SEN education since one of its primary objectives is to enable learners to use language functionally to engage in social interaction. Paradoxically, students with SENs typically face challenges in communication and self-management. How teachers will scaffold meaningful communication is an important aspect of the current TBLT framework.

(f) Theoretical and/or Conceptual Framework of the Project

The 5-dimensional TBLT framework underpinning this study's training, practising, and validation processes is based on the idea proposed in Chan (2022), which champions five key facets for

consideration for adapting TBLT in SEN contexts, and subsequently in Chan (2023). These facets include (a) person-environment reciprocity, (b) acting-perceiving processes, (c) three-world elements, (d) tools and affordances, and (e) specialist advice. As shown in the first three terms, the theoretical framework places a heavy emphasis on learners' internal perception of meaningfulness, on the goal of acting, and on the context in which learning activities take place. Through SCOLAR's support, these theoretical ideals can be translated into practice.

When applying them in the actual study (2021-2024), the first three concepts are combined into what I call a 'deep-dive' context in the remainder of this report. The 'acting and perceiving' processes (i.e., learners must take actions in order to perceive opportunities, and the perceived opportunities in turn inform output) further become 'task input' and 'task output' in the framework. With the transparent language, teachers and raters will be able to design tasks and understand evaluation purposes more effectively. The remaining two facets are preserved in the 5-dimensional framework. Altogether, the five aspects of TBLT implementation are: context, input, output, dialogical mediation and assistive. The verification of the 5 dimensions provides a tool for designing TBLT tasks and evaluating task-based lessons.

(g) Methodology

Lessons observed: The main research method of this study is class observations, based on the 5-dimensional observation scheme. This approach is psychometrically-oriented, designed to tap into the raters' interpretation of TBLT. As such, it is assessing truth through the lenes of rater judgement. The questions focused on (a) deep dive, (b) input, (c) output, (d) thinking-together, and (e) assistive techniques.

(h) Data collection and analysis

The number of lessons observed totalled 91, spanning 3,110 mins. The great majority of the lessons, 85, were observed by three observers to ensure maximum impartiality, while six lessons were observed by two observers. They added up to 237 times of observations.

Raters: They were members presented to SCOLAR, including the principal investigator, the research assistant (with a Ph.D), university language instructors (3), current and retired school principals (3), as well as a front-line educator with 20 years of experience (1). These observers were grouped in diverse combinations and assigned to rate the lessons using the following instrument, ensuring a diversity of perspectives. A rater could observe on-site or watch a video recording.

Instrument: After Chan (2022) published with a deeper understanding of literature reached, the final instrument differed from the original proposal. The questions re-categorised to highlight the importance of input, output and collaborative thinking to promote engagement (Original 5 aspects: story, undertaking tasks, perceiving meanings, using tools, practising skills). The five-aspect instrument consists of 46 items: 23 scalar questions on a Likert scale, 15 open-ended questions, and 3 multiple-choice questions. The scalar questions enabled the generation of aggregate and statistical data. The open-ended questions captured explanations which could be thematically coded via MAXQDA. The multiple-choice questions allowed teachers to choose from predefined options, limiting ambiguity that can arise from open-ended questions.

Table 2. Basic instrument for analysis – 5-Dimensional TBLT framework

Dimensions	Scalar questions (Generate aggregate and statistical data)	Open-ended questions
Summative questions (Independent of Dimensions 1-5)	How 'task-based' was this lesson? (Are there things to accomplish? Is there a mission?) (1…5) Was this an effectively conducted lesson?	1. What were some good practices in this lesson? 2. What were some poor practices in this lesson? 3. Any explanations for the strong/weak points?

		4. How "task-based" is this lesson? (Are there things to accomplish? Is there a mission?) 5. Why/Why not 'task-based'? 6. Was there a moment when communication failed (meaning is not comprehended)? How did the teacher solve the problem? 7. Was this an effectively conducted lesson?
DIM 1 (Deep dive)	D1.1 The context/theme has social/community relevance. D1.2 The context is real-life-based. D1.3 There are things to accomplish. D1.4 The context can be comprehended by students with ID. D1.5 There is intellectual stimulation / reaching the task goal involves concept development.	D1.6 What is your opinion of the task context?
Dim. 2 (Input)	D2.1 Linguistic input is well-defined in a task. D2.2 Linguistic input is gauged at the learners' level. D2.3a Task design encourages concept development. D2.4 There is a variety of activities in this lesson (e.g. listening, reading, role-play, information-gap.) D2.5 Tasks are appropriately ordered/sequenced. D2.6 Task arouses experiential knowledge. D2.7 Task arouses interpersonal knowledge / interpretation communication skills.	D2.3b Examples of concept development? D2.8 Do you have any comments or observe major problems with the task input?
Dim. 3 (Task output)	D3.1 Teacher/Task encourages students to respond (verbally, through gestures, moving, etc.). D3.2 Teacher/task encourages students to draw on prior knowledge. D3.3 Teacher/task encourages production of everyday/social knowledge. D3.4 Teacher/task evoke affective meanings.	D3.5 Do you have any comments or observe major issues with the task output?

Dim. 4 (Dialogical mediation and tools)	D4.1 Teacher's language is able to lead students to the theme/context. D4.2 Teacher makes explicit the concept(s) through language. D4.3 Teacher helps learners to try and try again via different examples. D4.4 Teacher demonstrates these classroom language: Wondering aloud (哈, 點解呢...), echoing students' response (無錯架...), accepting paraphrasing (係喇, 咁即係...) and modelling ways of thinking (我覺得呢, 如果我係呢, 我會...).	D4.5c What are your comments about the teacher's dialogic support and discursive strategies? Quality? Superfluous? Procedural? Distracting attention?)
Dim. 5 (Assistive techniques)	D5.1a Assistance offered is multisensory 1) Visual 2) Aural 3) Kinesthetic 4) PowerPoint clarity, 5) E-resources	D5.1b Use of tangible teaching tools, realia and other aids ? D5.2 Did you observe these in this lesson? Raising phonological awareness, (Spelling / sound-letter decoding) Repetitions (Words / syllables / sounds), Specially designed worksheets, Assistive technologies (Widgit symbols, writing aids, etc.)
Last questions		• Can you identify any turning points? • If you gave this teacher advice before, does he/she follow it? • After observing this lesson, what advice would you give to this teacher?

Main research questions (as stated in the proposal):

1. What is the internal reliability of the five-dimensional observation tool?
2. Can the five-dimensional TBL framework be suitable for guiding task designs for students with mild ID, for assessing teaching and learning? What is the difference between phase 1 and phase 2?
3. How might the 'training-practising-validation' phases inform areas for improvement – which dimension(s) of the new TBL framework needs fine-tuning? Which dimension(s) requires further development?

4. To what extent can the new TBL framework affect SEN students' learning? (In the proposal)

(i) Results and discussion (Quantitative data)

Table 3. Overview of the results presented in this report.

	Statistical analyses	Tools	Findings
Quantitative	Internal consistency	Cronbach's alpha coefficient	The internal consistency of the instrument and its various dimension is high.
	Interrater reliability	Interclass correlation	The raters did not diverge significantly in their rating of TBLT nature
	Improvement from Unit 1 to Unit 2	Means and standards deviations	The teachers' TBLT lessons and teaching skills can be quantified. Individual teachers have individual gains and teaching styles.
	Predicting the reasons for TBLT	Multiple regression	Dimensions 1, 2 and 4 are significant reasons for why a lesson is task-based.
	Refining the instrument	Factor analysis	The instrument is further refined to make it more robust
Qualitative	Teachers' insights	Interviews and reflections and thematic analysis	Learnt about the teachers' individual preferences. Various dimensions of the framework have been verified.

By answering these questions, this study examines whether existing task-based principles can be extended to the teaching of learners with mild IDs. It also investigates how TBLT might need to be redefined for this specific learner group.

Preliminary: Checking internal consistency

In a survey that relies on psychometric assessment of variables (i.e., raters' evaluations of question items), the first issue I must address is internal consistency reliability. This concept determines

whether an instrument is reliable for measuring what it claims - "groups of items that are thought to measure different aspects of the same concept" (Litwin, 2013, p.14)". The Cronbach's alpha coefficient (a measure of scale reliability) of the 28-scalar questions was 0.93. When the five items related to the 'Assistance' dimension were excluded, the reliability of the remaining 23 items is 0.92. The alpha value of dimension 1 was .830, dimension 2 .797, dimension 3.729, dimension 4.712 (after deleting question D4.4), and assistance .670. These figures signal that the instrument was 'adequate' according to Taber (2018). The alpha coefficient variations indicate that parts of the instrument can be fine-tuned, and a revised instrument will also be presented.

Table. Alpha coefficients by dimension

Cronbach Alpha	All items (27 items)	TBLT, EFF, D1 to D4 (22 items)	DIM 1 (5 items)	DIM 2 (7 items)	DIM 3 (4 items)	DIM 4 (4 items)	Visual Aural Kinetic PPT clarity e-Resource (5 items of Assistance)
0.64–0.85 Adequate (Taber, 2018)	0.933 Delete 4.4	0.927 Delete4.4	0.830	0.797	0.729	0.712 Delete 4.4	0.670

Table 4. Alpha coefficients by question

		Item-Total Statistics	
		Corrected Item-Total Correlation	Cronbach's Alpha of the dimension if Item Deleted
DIM 1 Deep Dive context (α= 0.830)	D1.1 (community)	0.670	0.784
	D1.2 (context)	0.723	0.769
	D1.3 (goal)	0.595	0.807
	D1.4 (difficulty)	0.568	0.812
	D1.5 (stimulation)	0.591	0.806
DIM 2 Input (α= 0.797)	D2.1 (well-definedness)	0.458	0.783
	D2.2 (appropriacy)	0.415	0.790
	D2.3 (concept develop)	0.655	0.746
	D2.4 (variety)	0.501	0.776
	D2.5 (sequence)	0.653	0.745
	D2.6 (experiential mean.)	0.572	0.762
	D2.7 (interpersonal mean.)	0.460	0.788
DIM 3 (α= 0.729)	D3.1 (verbal/nonverbal output)	0.423	0.72

	D3.2 (prior knowledge)	0.598	0.619
	D3.3 (everyday knowledge)	0.648	0.588
	D3.4 (emotions/affect)	0.427	0.724
DIM 4 (α= 0.709)	D4.1 (storytelling; intersubjective)	0.420	0.694
	D4.2 (make concepts explicit)	0.709	0.547
	D4.3 (exemplify)	0.579	0.606
	D4.5b (multisensory)	0.386	0.755
DIM 4 Assistance (α= 0.670)	Visual	0.569	0.583
	Aural	0.239	0.696
	Kinetic	0.408	0.632
	PPT clarity	0.572	0.571
	e-Resource	0.467	0.602

According to the table above, all items within Dimensions 1 to 3 appear to be integral to the measurement of the target Dimensions; removal of any item would lower the reliability of that dimension. Item D4.5 (Assistance offered is multisensory) in Dimension 4 had the lowest item-total correlation. One reason is that the question attempts to measure several aspects, and a rater may interpret 'multisensory' differently. Removal of the item would raise the level of reliability of Dimension 4. Similarly, item Aural Assistance had the lowest item-total correlation; removing it would raise the reliability of the dimension of Assistance.

As this is the first time that TBLT is focusing on LSEL (language, social, and emotional learning), items related to it suggest a connection between task-based learning and LSEL.

Preliminary: Checking Interrater Reliability

Although all the raters received briefings prior to the task, another preliminary step was to investigate the consistency of their judgement, that is, the degree to which different raters agreed on particular variables (Litwin, 1995). Table 5 shows the teachers observed: [redacted]

[redacted]

Table 5. Interclass correlation for each teacher's observed lessons

Teachers observed	ICC	95% confidence intervals (CI)	Interpretation (Eltayar et al., 2022)	Interpretation (Cicchetti & Sparrow, 1981)
[redacted]	0.91	0.895 - 0.925	very good	excellent
	0.92	0.896 - 0.932	very good	excellent
	0.94	0.943 - 0.955	very good	excellent
	0.93	0.924 - 0.939	very good	excellent
	0.94	0.923 – 0.947	very good	excellent
	0.92	0.891 - 0.946	very good	excellent
	0.92	0.855 - 0.941	very good	excellent
	0.92	0.901 - 0.924	very good	excellent

The interclass correlation (ICC) method was a preferred method to measure interrater reliability due to its suitability for analysing continuous data. The ICCs were calculated using a two-way mixed effect model, which was appropriate when the same team of raters were observing multiple teachers. A two-way mixed method assumes that the raters were a fixed effect (the same raters are used across all measurements), while the teachers being rated were a random effect (the findings from them can be generalised to a wider population). In this study, the ICC for inter-rater reliability ranges from 0.91 to 0.94. The raters' ratings were highly consistent. This could be explained by the fact that wording of the questions did not lead to disparate interpretations.

Teacher's individual growth trajectories in Unit 1 and Unit 2

One crucial question asked by this study was 'Can the five-dimensional TBL framework be suitable for guiding task designs for students with mild ID?' After the preliminary analyses, the researcher began to investigate the crucial function of the instrument in assessing the SEN teachers' ability to connect TBLT and LSEL (language, social and emotional learning).

Recall that the instrument (Table 2) started with two summative questions: 'How task-based was this lesson?' ('How TBLT') and 'How effective was this lesson?' ('How effective). A typical lesson was evaluated by three raters, which produced the results below. For a detailed analysis, it is important to look at the mean and standard deviation (SD) for each dimension.

Table 6. Teachers' performance in Unit 1 (collaborating with researchers) and Unit 2 (working independently)

			Summative		Dimensions 1-5					Over all
			How TBLT?	How Effective?	Context	Input	Output	Collabt hink.	Assistance	
All teachers	U1	M (SD)	3.49 (0.929)	3.45 (0.944)	3.89 (0.634)	3.50 (0.643)	3.22 (0.822)	3.55 (0.879)	3.54 (0.738)	3.52
	U2	M	3.70 (0.905)	3.50 (0.825)	3.87 (0.615)	3.55 (0.593)	3.33 (0.728)	3.65 (0.659)	3.71 (0.722)	3.62
	Both	M	3.59 (0.922)	3.48 (0.890)	3.88 (0.624)	3.52 (0.619)	3.27 (0.780)	3.60 (0.785)	3.62 (0.734)	3.56
					↓0.01 Little change	↑0.05	↑0.11	↑0.10	↑0.17	

- In Unit 1 (collaborating with the researchers), TBLT score was 3.49, Lesson Effectiveness was 3.45, with means ranging from 3.22 to 3.89 for Dimensions 1 through 4, and a mean of 3.55 for Dimension 5 ASSISTANCE.
- In Unit 2 (working independently), TBLT rose to 3.70, Lesson Effectiveness was at 3.50, DIMs ranging from 3.22 to 3.87, and ASSISTANCE at 3.65. Unit 2 consistently shows higher means across all dimensions, suggesting more effective teaching practices.
- Lesson observers consistently rated Dimension 1 context highly, indicating the significance of this aspect, they praised teachers' efforts in Dimension 5 (visuals, audio input), followed by

Dimension 4 collaborative thinking. In adapting TBLT for LwSENs, Dimension 1 and Dimension 4 must meet the needs of the learners.

[redacted]	U1	M (SD)	2.93 (0.799)	3.27 (0.884)	3.52 (0.827)	3.27 (0.661)	2.85 (0.731)	3.18 (0.884)	3.42 (0.763)	3.21 (--)
	U2	M	3.19 (1.109)	3.44 (0.814)	3.64 (0.670)	3.43 (0.309)	2.95 (0.672)	3.61 (0.632)	3.91 (0.484)	3.45 (--)
	Both	M	3.06 (0.964)	3.35 (0.839)	3.58 (0.740)	3.35 (0.508)	2.90 (0.691)	3.40 (0.782)	3.67 (0.673)	3.33 (--)

↑Analysis: [redacted] shows improvement from Unit 1 to Unit 2, with TBLT increasing from 2.93 to 3.19 and Lesson Effectiveness from 3.27 to 3.44. DIM1 to DIM4 and ASSISTANCE also show increases, suggesting improved TBLT implementation.

[redacted]	U1	M (SD)	2.70 (0.949)	2.60 (0.843)	3.32 (0.675)	2.89 (0.819)	2.93 (0.746)	2.95 (1.218)	2.84 (1.103)	2.79 (--)
	U2	M	3.83 (0.857)	3.50 (0.857)	3.79 (0.684)	3.45 (0.651)	3.22 (0.737)	3.43 (0.751)	3.39 (0.649)	3.52 (--)
	Both	M	3.43 (1.034)	3.18 (0.945)	3.62 (0.706)	3.25 (0.753)	3.12 (0.741)	3.26 (0.951)	3.19 (0.861)	3.29 (--)

↑Analysis: [redacted] exhibits a significant rise in TBLT from 2.70 to 3.83 and Lesson Effectiveness from 2.60 to 3.50, with similar improvements across DIMs and ASSISTANCE.

[redacted]	U1	M (SD)	3.81 (0.877)	3.81 (0.967)	4.06 (0.538)	3.65 (0.635)	3.42 (0.736)	3.78 (0.710)	3.72 (0.624)	3.75 (--)
	U2	M	3.90 (0.845)	3.77 (0.728)	4.09 (0.495)	3.85 (0.568)	3.63 (0.694)	3.83 (0.634)	4.02 (0.588)	3.87 (--)
	Both	M	3.85 (0.857)	3.79 (0.862)	4.07 (0.515)	3.74 (0.609)	3.51 (0.720)	3.80 (0.672)	3.85 (0.622)	3.80 (--)

↑Analysis: [redacted] maintains high performance in both units, with slight increases from UNIT 1 to UNIT 2, such as TBLT rising from 3.81 to 3.90 and Lesson Effectiveness from 3.81 to 3.77. DIMs and ASSISTANCE also show consistent high ratings.

[redacted]	U1	M (SD)	3.62 (0.758)	3.38 (0.828)	3.95 (0.612)	3.53 (0.512)	3.05 (0.944)	3.48 (0.861)	3.49 (0.861)	3.5 (--)
	U2	M	3.67 (0.844)	3.37 (0.809)	3.83 (0.582)	3.46 (0.623)	3.30 (0.831)	3.63 (0.625)	3.42 (0.947)	3.53 (--)
	Both	M	3.64 (0.792)	3.37 (0.813)	3.89 (0.597)	3.50 (0.561)	3.16 (0.898)	3.54 (0.762)	3.46 (0.804)	3.51 (--)

↑[redacted] shows stable performance with slight increases in UNIT 2, with TBLT moving from 3.62 to 3.67 and Lesson Effectiveness from 3.38 to 3.37.

[redacted]	U1	M (SD)	3.33 (1.073)	3.83 (0.389)	4.12 (0.493)	3.75 (0.581)	3.75 (0.707)	4.02 (0.635)	3.87 (0.394)	3.81 (--)
	U2	M	3.60 (0.986)	3.00 (0.894)	3.47 (0.524)	3.22 (0.583)	3.23 (0.395)	3.30 (0.640)	3.73 (0.508)	3.41 (--)
	Both	M	3.39 (0.941)	3.43 (0.788)	3.81 (0.595)	3.50 (0.630)	3.50 (0.626)	3.67 (0.725)	3.80 (0.447)	3.59 (--)

↑ [redacted] demonstrates an increase in performance from UNIT 1 to UNIT 2, with TBLT increasing from 3.33 to 3.60. However, her Lesson Effectiveness decreased from 3.83 to 3.00. DIMs and ASSISTANCE also reflect this trend.

[redacted]	1	M (SD)	3.17 (1.169)	3.00 (0.632)	3.77 (0.543)	3.38 (0.674)	3.29 (0.534)	3.54 (0.534)	3.80 (0.400)	3.42 (--)

↑[redacted] only taught 2 lessons in Unit 1. She demonstrated a mean TBLT score of 3.17 and a Lesson Effectiveness (EFF) score of 3.00. Her performance across the dimensions showed a mean of 3.77 for DIM1, 3.38 for DIM2, 3.29 for DIM3, and 3.54 for DIM4, with an ASSISTANCE mean of 3.80.

[redacted] (withdrew after 3 lessons)	1	M (SD)	3.5 (1.049)	2.33 (1.033)	3.50 (0.486)	2.98 (0.654)	2.71 (0.557)	2.46 (0.641)	2.50 (0.415)	2.85 (--)

↑ [redacted] also taught 3 lessons in Unit 1 and withdrew from the study. [redacted] (CS) had a mean TBLT score of 3.52 and a Lesson Effectiveness (EFF) score of 2.33. His scores across the dimensions were 3.50 for DIM1, 2.98 for DIM2, 2.71 for DIM3, and 2.46 for DIM4, with an ASSISTANCE mean of 2.50.

[redacted]	U1	M (SD)	3.89 (0.782)	4.11 (0.782)	4.22 (0.273)	3.87 (0.281)	3.64 (0.697)	4.33 (0.354)	4.07 (0.374)	4.02 (--)
	U2	M	4.13 (0.835)	3.88 (0.835)	4.45 (0.397)	3.70 (0.542)	3.41 (0.582)	4.13 (0.354)	3.93 (0.320)	3.94 (--)
	Both	M	4.01 (0.808)	3.99 (0.808)	4.34 (0.335)	3.78 (0.412)	3.52 (0.639)	4.23 (0.354)	4.00 (0.347)	3.98 (--)

↑ [redacted] (WB) co-taught all lessons in the study, an arrangement not commonly seen in SEN schools. These teaching partners demonstrated high performance across both units, with TBLT score increasing from 3.89 to 4.13 and Lesson Effectiveness from 4.11 to 3.88. DIMs and ASSISTANCE also show high scores, indicating effective teaching strategies.

Discussion:

Analysing teachers' profiles shows that TBLT implementation and lesson effectiveness can overlap but be distinct concepts. However, TBLT generated more intersubjective attention to the teaching content, as well as more interesting and localised elements. Teachers can experience different directions of growth. Some were evidently more adept at dialogic mediation and some may want to focus on assistive techniques.

Multiple regression: Which dimensions predict a task-based lesson?

If the five dimensions (context, input, output, dialogic mediation, assistance) are assumed to assess a teacher's TBLT implementation, they should predict the teacher's TBLT score (At the top of the instrument, we asked raters to assign a score). This is the purpose of using multiple regression for this analysis: whether Dimensions 1-5 accurately predict a lesson's task-basedness. Validating these dimensions is crucial for clarifying the nature of TBLT in SEN contexts, helping us understand what lesson raters expect to observe and factors that determine (or not) a task-based lesson. Specifically,

Dimensions 1 to 4 as well as assistive techniques were predictors, while perceived TBLT implementation was the dependent variable (Keith, 2015). Recall that TBLTness was a summative score given by the raters in the instrument.

Extent of TBLT											
Time	R^2	f^2	F	*df*	*p*	IV	Beta	*t*	*p.*	VIF	Predict
UNIT 1	**0.435**	0.770	19.400	5, 126	<.001	DIM1	0.466	3.869	**<.0001**	3.239	Yes
						DIM2	0.508	4.068	**<.0001**	3.479	Yes
						DIM3	0.043	0.476	0.635	1.792	
						DIM4	-0.268	-2.050	**0.042**	3.820	Yes
						ASSISTANCE	-0.144	-1.175	0.242	3.344	
UNIT 2	**0.527**	1.114	23.865	5, 107	<.001	DIM1	0.459	4.421	**<.0001**	2.436	Yes
						DIM2	0.265	2.046	**0.043**	3.802	Yes
						DIM3	0.239	2.471	**0.015**	2.114	Yes
						DIM4	-0.029	-0.274	0.785	2.597	
						ASSISTANCE	-0.207	-2.023	**0.046**	2.364	Yes
OVERALL	**0.458**	0.845	40.406	5, 239	<.001	DIM1	0.444	5.624	**<.0001**	2.745	Yes
						DIM2	0.426	4.758	**<.0001**	3.542	Yes
						DIM3	0.110	1.699	0.091	1.850	
						DIM4	-0.187	-2.219	**0.027**	3.115	Yes
						ASSISTANCE	-0.138	-1.753	0.081	2.741	

↑ **Outcome 1: Task-based nature:**

Trend:

In **Unit 1 (collaborating with researchers)**, the R^2 (which measures the fit between the data and the statistical model; a higher R^2 indicates a better fit) for TBLT is 0.435, indicating that the whole set of predictors explains 43.5% of the variance. The significant predictors include DIM1 (Beta = 0.466, p < .0001), DIM2 (Beta = 0.508, p < .0001), and DIM4 (Beta = -0.268, p = 0.042). DIM3 and Assistance are not significant predictors. In **Unit 2 (without independently)**, the R^2 for TBLT increases to 0.527, indicating that the predictors explain 52.7% of the variance. Significant reasons for task-basedness are DIM1 (Beta = 0.459, p < .0001), DIM2 (Beta = 0.265, p = 0.043), and DIM3 (Beta = 0.239, p = 0.015). Assistance was also a significant predictor. When considering all lessons in Units 1 and 2, the R^2 is **0.458**, meaning that nearly half of what happens (45.8%) in a classroom can be attributed to task-basedness. The significant reasons for

task-basedness are DIM1 (Deep dive context, Beta = 0.444, p < .0001), DIM2 (Input, Beta = 0.426, p < .0001), and DIM4 (Teacher's dialogic mediation, Beta = -0.187, p = 0.027).

Discussion:

What does it mean if these reasons account for only 45.8% of the model? When evaluating a teacher's holistic performance, the raters might take into account other factors. At times, the instrument cannot assess all aspects of a prototypical TBLT lesson. There may be aspects of TBLT that are prototypical - which define what TBLT is - that cannot be observed (for example, non-verbal students cannot negotiate meaning).

Lesson Effectiveness											
Time	R^2	f^2	F	*df*	*p*	IV	Beta	*t*	*p.*	VIF	Predict
UNIT 1	**0.616**	1.604	40.500	5, 126	<.001	DIM1	-0.044	-0.446	0.656	3.239	
						DIM2	0.473	4.599	**<.0001**	3.479	Yes
						DIM3	0.057	0.771	0.442	1.792	
						DIM4	0.288	2.675	**0.008**	3.820	Yes
						ASSISTANCE	0.080	0.796	0.428	3.344	
UNIT 2	**0.623**	1.652	35.400	5, 107	<.001	DIM1	0.126	1.358	0.177	2.436	
						DIM2	0.543	4.692	**<.0001**	3.802	Yes
						DIM3	0.023	0.262	0.794	2.114	
						DIM4	0.131	1.368	0.174	2.597	
						ASSISTANCE	0.043	0.468	0.641	2.364	
OVERALL	**0.615**	1.597	76.403	5, 239	<.001	DIM1	0.046	0.697	0.487	2.745	
						DIM2	0.494	6.546	**<.0001**	3.542	Yes
						DIM3	0.052	0.953	0.342	1.850	
						DIM4	0.223	3.148	**0.002**	3.115	Yes
						ASSISTANCE	0.045	0.682	0.496	2.741	

↑ Outcome 2: Lesson effectiveness

Trend:

Non-task-based researchers have their own beliefs about what constitutes effective teaching (examples are given by Hall 2018; Sato & Loewen, Laurillard 2012) without having to adopting TBLT; a task-based lesson that is not effectively delivered cannot produce good learning. It would be interesting to compare whether the raters would rate a lesson the same way and the same reasons would predict lesson effectiveness. In **Unit 1**, the R^2 for lesson effectiveness is

0.616, indicating that the predictors explain 61.6% of the variance. Significant predictors were DIM2 (Beta = 0.473, $p < .0001$) and DIM4 (Beta = 0.288, $p = 0.008$). DIM1, DIM3, and Assistance were not significant predictors. In **Unit 2**, the R^2 for lesson effectiveness was 0.623, explaining 62.3% of the variance. The significant predictor is DIM2 (Beta = 0.543, $p < .0001$). DIM1, DIM3, DIM4, and Assistance are not significant predictors. Also, the R^2 had increased from Unit 1 to Unit 2.

Discussion:

When considering all lessons in Units 1 and 2, the R^2 for lesson effectiveness was 0.615, explaining 61.5% of the variance. Significant predictors include DIM2 (Beta = 0.494, $p < .0001$) and DIM4 (Beta = 0.223, $p = 0.002$). The outcome suggests the raters ignored context and focused on well-definedness of the teaching points, concepts (DIM2) and the teachers' scaffolding (DIM4). The inclusion of Dimension 1 as part of the framework orients expectations' attention towards Context. The failure to achieve a higher R^2 may be due to the inability to assess some TBLT principles.

Specialist techniques:

The instrument also asked raters to observe any specialist techniques used. Regarding DIM5, Repetition was the most frequently used technique across all units, with 33.0% in UNIT 1, 33.5% in UNIT 2, and 32.6% overall. Spelling/Decoding (SD) and Phonological Awareness (PA) also had significant percentages, indicating their importance in teaching practices. Only a few worksheet designs (12.8%) were considered thoughtful for students with SENs. Many worksheets that we observed were not connected or they were loose pages without a sense of coherence. IT, which is a generic tool nowadays, is the most commonly used tool in the classroom (but not specialist techniques).

Factor Analysis – Can the 5-D framework be refined?

After investigating the nature of task-based teaching in SEN classrooms, the next step involved refining the 5-D framework for task design. We used Partial Least Squares Structural Equation Modeling (PLS-SEM) to increase the reliability of the constructs (dimensions) measured by the framework.

Procedure:

PLS-SEM was conducted with SmartPLS 4 (https://www.smartpls.com/downloads/). The SEM-PLS model was initiated after the removal of D4.5 and D4.5 Aural, due to their low item-total correlations. In the first step, the factor loadings of D2.1 (r.= 0.595), D2.2 (r.= 0.569), and D4.5 Kinesthetic (r.= 0.598) were found to be below 0.6. These items were removed in the second round. Additionally, the Average Variance Extracted (AVE) for Dimension 2 'Input' was 0.458, below 0.5, indicating the composite reliability for this dimension was unsatisfactory. Furthermore, the Context dimension failed to meet the Fornell-Larcker Criterion as the square root of the AVE for Context (0.772) was lower than its correlation with 'Input' (0.775), suggesting low discriminant validity. Similarly, the square root of the AVE for Input (0.677) was lower than its correlation with Mediation (0.775). It was anticipated that deleting D2.1 and D2.2 would resolve the reliability and validity issues. After the deletion of D2.1, D2.2, and D4.5 Kinesthetic in the second round, the factor loadings of all remaining items reached acceptable levels.

Function:

The PLS-SEM method is particularly useful in identifying questionnaire items that display suboptimal loadings. For example, D4.5b and D4.5 (Aural) had lower item-total correlations, and poor loadings refer to variables that do not seem to measure underlying or assumed

constructs. For practical application, the PLS-SEM refined the instrument's questions, reducing its 21 scalar questions to 15 questions. The refinement process not only improves the precision of the existing instrument, but also provides a clearer, more concise measure of TBLT as a concept for task design and evaluation of teaching.

With PLS-SEM, the questionnaire items with lower loadings leading to poor reliability and validity were identified and deleted.

Table 7. A refined instrument for students with mild ID or SENs

Dimension	**Indicator**	**Convergent Validity**		**Internal Consistency**		**Discriminant Validity**
		Loadings	**AVE**	**Cronbach Alpha (rho_c)**	**Composite Reliability (CR)**	**Fornell-Larcker criterion**
		≥.70 .40 to .69 if CR≥.70	**≥.50**	**.60 to .90**	**.60 to .90**	$\sqrt{AVE_A} > Corr_{A\&other\ constructs}$
DIM1 Context	D1.1	0.790	0.632	0.873	0.807	Yes
	D1.2 The context is real-life-based.	0.825				
	D1.4 The context can be comprehended by students with ID.	0.777				
	D1.5 There is intellectual stimulation.	0.788				
	~~D1.3~~ (removed)					
DIM2 Task Input	D2.3 Task design encourages concept development.	0.816	0.554	0.860	0.797	Yes
	D2.4	0.644				

	There is a variety of activities in this lesson.					
	D2.5 Tasks are appropriately ordered/sequenced.	0.808				
	D2.6 Task arouses experiential knowledge.	0.777				
	D2.7	0.657				
	~~D2.1~~					
	~~D2.2~~					
DIM3 Task Output	D3.1 Task encourages students to respond.	0.689	0.554	0.831	0.728	Yes
	D3.2 Task encourages students to draw on prior knowledge	0.782				
	D3.3 Task encourages production of everyday/social knowledge	0.841				
	D3.4 Teacher/task evoke affective meanings	0.651				
DIM4 Dialogic mediation	D4.1 Teacher's language is able to lead students to the theme/context	0.773	0.678	0.863	0.760	Yes
	D4.2 Teacher makes explicit the concept(s) through language.	0.886				
	D4.3 Teacher helps learners to try and try again via different examples.	0.807				

	~~D4.4~~					
DIM5 Assistance	D4.5 PPT	0.898	0.704	0.876	0.786	Yes
	D4.5 E-Resource	0.726				
	D4.5 Visual	0.884				
	~~Aural~~					
	~~kinesthetic~~					
Overall	Context	0.862	0.678	0.913	0.880	Not Applicable
	Input	0.877				
	Output	0.800				
	Mediation	0.854				
	Assistance	0.714				

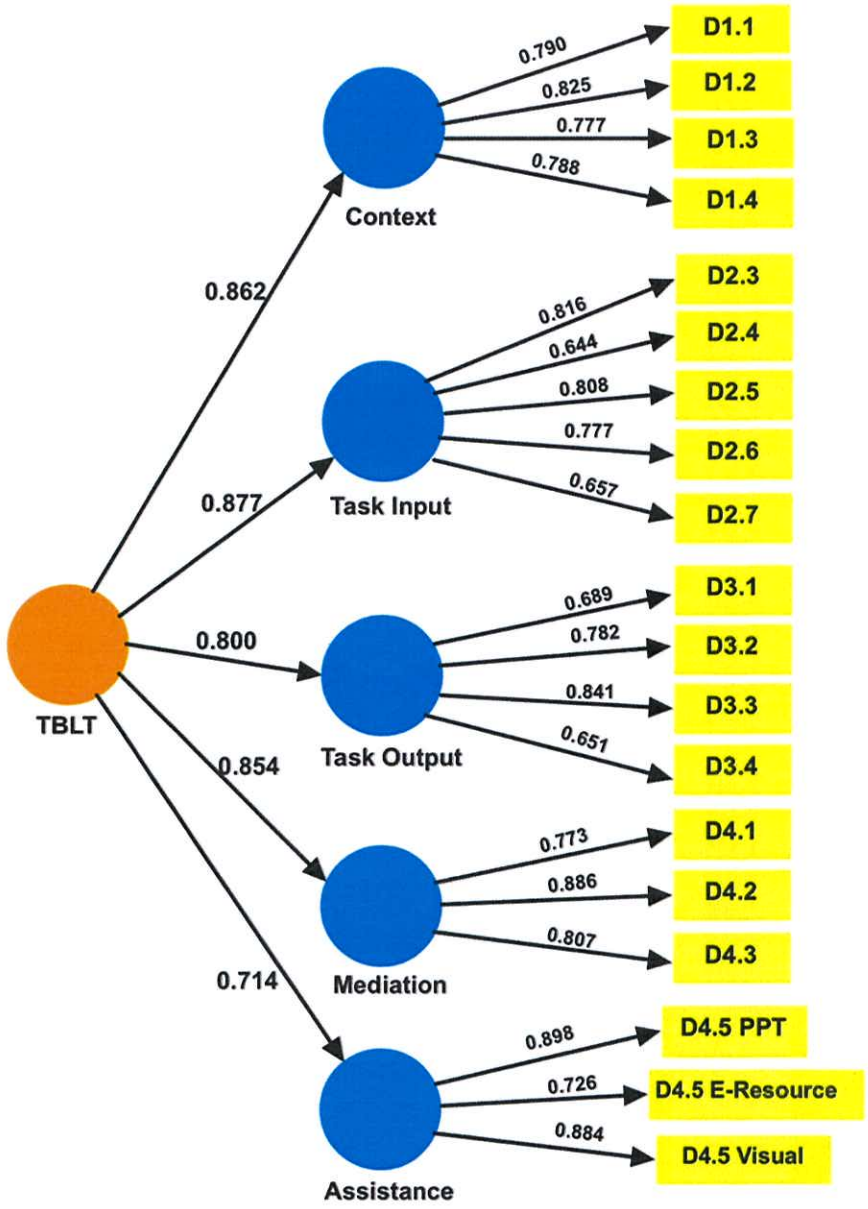


Figure. Correlations of the different factors in the 5-factor model of the TBLT

(i) Results and Discussion (Qualitative data)

Qualitative data are important for complementing the interpretation of the quantitative data and the validation of the 5-dimensional principles. The interviews were particularly informative for illustrating the teachers' insights after the practising stage. To this end. 7 teachers' pre- and post-interviews were transcribed and analysed, with a total of 158,259 words.

Before Training	After Practising	Discussion

[redacted]

(Significance for this study)

[redacted]'s TBLT score was average. She received a mark of 3.43/5 for the question 'How task-based is this lesson'.

Before the training stage, she **was able to name many interesting tasks design skills**. She mentioned a wide range of TBLT principles, indicating her **loose understanding of TBLT or an inability to grasp the essential task ingredients** she had to focus on.

After the training stage, she also provided a wide range of learnt principles of TBLT. Interestingly, she **mentioned the importance of decoding strategies** (i.e., helping students grasp phonemic concepts during tasks) **although decoding is not a fundamental principle of TBLT.** (Dimension 5)

She admitted to having difficulty sequencing and linking tasks. This **indicates that the ability to plan long-term and sustainable content for teaching remains a main obstacle for task designers**. (Dimension 1)

After working with [redacted], the team realised that she failed to pay attention to the intricate skills of students with ID and **focused only on the superficial task designs**. She initially wanted Mr Beans to train students' sense-making skills, but **this focus was lost or not maintained**. **Grammar drills became the focus of the lesson**. (Dimension 1)

Such reflections were a result of her lack of detail-orientedness or using more effective scaffolding language. (Dimension 4). The project suggests that **planning in detail is crucial for a task's success.**

[redacted]

(Significance for this study)

[redacted]'s TBLT score was lower than the group's average. She received a mark of 3.060/5 for the question 'How task-based is this lesson' based on 10 lessons observed by three observers each (33 times).

However, **she learns fast and listens to advice quickly**. There were movements that she **felt touched and useful when she saw that her student could achieve task goals**. However she prefers more **flexibility in teaching approaches and more guidance from teacher educators when it comes to task design.** (Dimension 1)

Her method of learning involves **deliberate use of TBLT** and listening to the project team's advice. She had difficulty in **designing the task context, defining task input, and planning concepts to develop among the students**. (Dimensions 1 & 2)

[redacted]'s classroom language was weak; she frequently used Cantonese and said things that distracted students from their work. She did a lot of things that were not task-based.

We can learn from [redacted]'s story that **teachers' language and questioning skills are fundamental to enact tasks and their contexts**. (Dimension 4)

[redacted]

(Significance for this study)

[redacted]'s TBLT score was in the middle of the group of teachers, with a mark of 3.394/5 for the question 'How task-based is this lesson?'

Before the Training stage, [redacted]'s understanding of TBLT was **incomplete and fragmented.** She associated it with music, movement, enjoyment, and collocation, but the essence of TBLT is the **promotion of meaning-making practices**, no matter what methods are used to promote it.

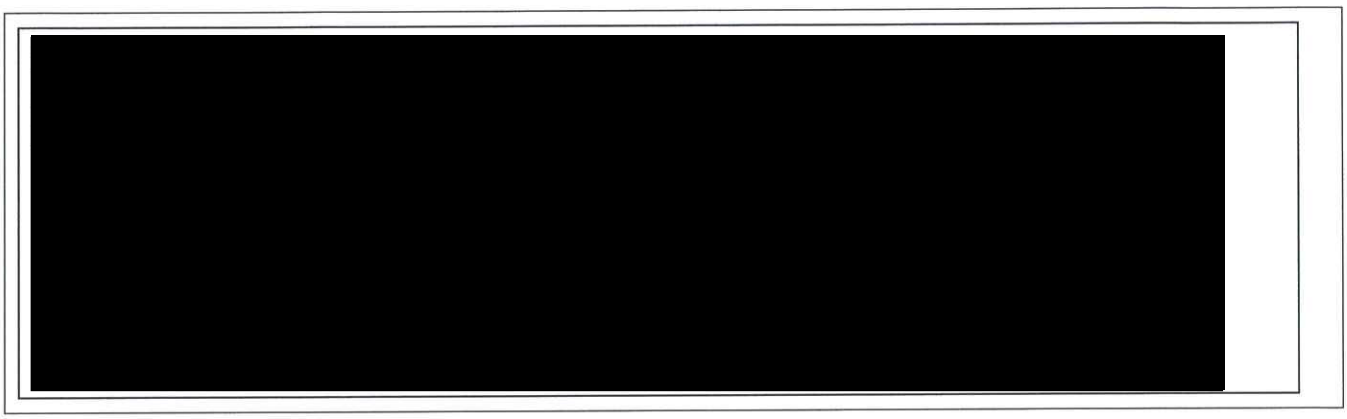
After practising, she realised that her TBLT lessons emphasised **developing students as sentient human beings**, (Dimensions 1 & 2) teaching them to be sensitive and aware (her teaching topic were five senses and emotions).

She became more aware of tasks' constraints and possibilities. She recognised that **designing thematic tasks while meeting 'school-prescribed learning targets' is indeed a challenge**. (Dimension 1)

Finally, she realised that she had to train students to **use different skills**, apart from listening. (Dimension 2)

We noticed that [redacted] **initially resisted TBLT, considering it too creative or too open to imagination.**

Through a collaborative approach to lesson planning, there was an increased adoption of creative elements through multimedia and videos (Dimension 5).



(Significance for this study)
[redacted] is the third strongest teacher among all teachers, with a mark of 3.644/5 for the question 'How task-based is this lesson?' after 22 lessons observed by raters.

Before the study, he received standard explanations of theories, but after it, he has developed a more **nuanced understanding of the possibilities of TBLT**, including **information gaps**, **incidental learning**, **practising the four skills**, and the **emotional significance** when designing tasks. (Dimensions 1 and 2). [redacted] asks many questions about TBLT theories in the post-study interview.

However, there is still much room for improvement in his **dialogic mediation skills** (Dimension 4). He did not give enough wait time or prompt them when eliciting answers.

[redacted]

(Significance for this study)

[redacted] ranked second among all teachers in adopting TBLT, with a mark of 3.855/5 after 22 lessons observed by three observers each.

Before the Training stage, [redacted] had a **clear idea** of how her teaching should **incorporate gamified elements and presented through a story**. (Dimension 1)

She is the only teacher who believes **critical thinking should be emphasised through doing tasks** in special schools. (Dimension 2). Her students' ability was relatively high.

During the Practising stage, her idea was tested, and she gained a nuanced understanding of TBLT's malleability. **She perceives that an imaginary task differs from a real-world task** (Dimension 1) **, and how both can foster students' reflective abilities**. (Dimension 2)

Besides language learning, she believes that TBLT is associated with conceptual development and higher-order thinking (e.g., via problem-solving activities). (Dimension 2) [redacted]'s insight is **prescient**.

[redacted]

(Significance for this study)

[redacted] co-taught all lessons. Their lesson observation mark was 4.01/5 based on 42 observations over 14 lessons. The pair ranked first in teaching skills and TBLT adoption.

Before the Training stage, [redacted]'s task schema was **rather general and standard**; she could name the key terms related to standard TBLT, but **her viewpoint was not very strong**.

After the practising stage, [redacted]'s **understanding became refined**, and she is particularly **concerned about the cognitive load brought by the task design**. She understands the role of task in the **strategic organisation of activities and its exploratory nature**. She intends to engage the students in an exploration journey together. (Dimension 1).

She demonstrates an ability to **transfer her skills to the next unit**, where she has envisioned how the different tasks might be. (Dimension 1)

After practising, [redacted] is more cautious about using tasks, because she is concerned about students' cognitive load. She was able to assess their pros and cons.

The team found this to be one of the **best teachers for providing dialogical mediation and scaffolding**, even though they did not mention it. **They acted out the tasks and asked many questions in order to invite the students into the task contexts**. This is one of the successful ingredients in adapting TBLT to SEN contexts. (Dimension 4)

[redacted]

(Significance for this study)

[redacted] co-taught all lessons. Their lesson observation mark was 4.01/5 based on 42 observations over 14 lessons.

Before the Training stage, [redacted] provided standard responses about TBLT. However, after practising, she was able to identify several different strategies, such as **'learning by doing' for concept development, sequential progression** (Dimension 2), and **teachers improvising and enacting a task** (Dimension 4).

From her reflection, [redacted] supports **task-'on-the-go', or process tasks**. Tasks allow students to see how scenarios pass and evolve as they are taking part in activities and receiving input (Dimension 1).

This notion of TBLT emphasises the fact that **tasks are not external static missions**, but **a reality that is happening** right before the student's eyes (Dimension 1). It is critical to understand this aspect when defining a task in SEN classrooms.

Her views after practicing echo that **teacher's language** is **an essential ingredient for building a joint identity and a shared viewpoint between the teacher and the students on the task at hand**.

(j) Conclusions and Recommendations

When assessing teachers' task-based strategies using the 5-dimensional TBLT framework, it is found that the teachers had increasingly adopted it and their understanding of the approach had been strengthened after the Practising stage. Their differing expectations of using tasks are characterised by their viewpoints - for example, some see TBLT as 'learning by doing' while others see it as a means of

promoting concept learning and sense-making. Varying degrees of adoption are referred to as 'variance'. One key finding of the study is that the 5-dimensional framework explains 45.8% of the variance in task-basedness. The same framework can be used to assess teaching effectiveness, when it explains 61.5% of the variance.

Based on close collaboration with 7 teachers across 6 schools, and the results from 273 times of observations (91 lessons) , this study has developed a practical, applicable 5-dimensional TBLT framework for teachers and assessors. Specially, the instrument can be used for evaluating as well as task planning. It is designed to set a ceiling for teachers to work towards.

This version of TBLT has emphasised creating an interesting context and teachers' dialogical skills to facilitate learning. Some aspects of it are not fully utilised, suggesting that there may be a limit on how much it can be used by teachers. After all, if a teacher lacks planning skills, they will find designing a deep dive context that promotes personal investment (Lambert , 2023) challenging. Similarly, without dialogical mediational skills or improvisational abilities, they cannot effectively lead learners into scenarios that are personalised, inviting, and call upon prior memory. They might only be able to deliver a textbook version of tasks.

Table 8. The following example illustrates how 'task' can be redefined.

Features of task-based learning (Ellis, 2009; Willis & Willis, 2007; Skehan, 1998)	TBLT extended (underlined)	Evidence
Primary focus on meaning	Focus on meaning <u>in a rich, emotionally-investing context a involves a deep dive into scenarios</u>	Dimension 1, which predicts TBLT
Real-world focus	Real-world scenarios <u>illustrated and mediated by lively language and visuals</u>	Dimension 1, which predicts TBLT Dimension 4, which predicts TBLT
Clearly defined outcomes other than the use of language	Clearly defined outcomes other than the use of language, <u>e.g., paralinguistic productions and non-verbal behaviour.</u>	Teachers' interviews

Some communication problems to solve (students use language to solve problems)	Some communication problems to solve with the help of pictures, clear illustrations and teachers' dialogic support.	Dimension 4, which predicts TBLT
Tasks should go beyond the practice of language for its own sake Richards, Platt and Weber (1985)	Tasks should go beyond the practice of language for its own sake to facilitate concept development (e.g., sense-making) and a sense of community	Dimension 1 of the framework
A task is a holistic activity which engages language use in order to achieve some non-linguistic outcome. Samuda and Bygate (2008)	A task is a holistic activity that involves joint attention and mutual discovery, which engages language use in order to achieve some non-linguistic outcome, such as meaning on an emotional, interpersonal, or experiential level. A task is a holistic activity that has elements of a project or journey.	Teachers' interviews; teachers' recorded lessons
A task that should support negotiation of form and content in a holistic activity	A task that should support negotiation of form and content is a holistic activity that involves teacher-student interaction and elicitation of forms. The designing of grammar judgement tasks, picture description tasks, and special worksheets were also helpful in negotiating meaning.	We have observed meaning negotiation in the lessons of [redacted]

References:

Boers, F., & Faez, F. (2023). Meta-analysis to estimate the relative effectiveness of TBLT programs: Are we there yet? *Language Teaching Research, 0*(0). DOI: 10.1177/13621688231167573

Bolton, N. (1977). *Concept formation* (1st ed. ed.). Oxford: Pergamon Press.

Bryfonski, L., & McKay, T. H. (2019). TBLT implementation and evaluation: A meta-analysis. *Language Teaching Research, 23*(5), 603-632. DOI: 10.1177/1362168817744389.

Chan, H. (2022). Adapting the task-based methodology for learners with intellectual disabilities: Five key facets for consideration. *International Journal of Applied Linguistics, 32*(3), 459-475. DOI: 10.1111/ijal.12440.

Chan, H. (2023). Ideating task-based designs for special education learners. TESOL Journal, 14(3), e711. DOI: 10.1002/tesj.711 .

Cicchetti, D. V & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86(2),127-137.

Cohen J. (1992). A power primer. Psychological bulletin, 112(1), 155–159. DOI: 10.1037//0033-2909.112.1.155

Costello, Anna B. & Jason Osborne (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment Research & Evaluation, 10(7). Available online: http://pareonline.net/getvn.asp?v=10&n=7.

Daloiso, M. (2017). *Supporting learners with dyslexia in the ELT classroom*. Oxford University Press.

Ellis, R. (2005). Principles of instructed language learning. *System, 33*(2), 209-224.

Eltayar, A. N., Aref, S. R., Khalifa, H. M., & Hammad, A. S. (2022). Do entrustment scales make a difference in the inter-rater reliability of the workplace-based assessment? Medical Education Online, 27(1), 2053401–2053401. DOI: 10.1080/10872981.2022.2053401

Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, Calif.: Sociology Press.

Godfroid, A., & Andringa, S. (2023). Uncovering sampling biases, advancing inclusivity, and rethinking theoretical accounts in second language acquisition: Introduction to the special issue SLA for All? *Language Learning, 73*(4), 981-1002. DOI: 10.1111/lang.12620

Hair, J., Black, W. C., Babin, B. J. & Anderson, R. E. (2010) Multivariate data analysis (7th ed.). Upper Saddle River, New Jersey: Pearson Educational International.

Hiver, P., & Wu, J. (2023). Engagement in TBLT. In C. Lambert, S. Aubrey, & G. Bui (Eds.), *The role of the learner in task-based language teaching: Theory and research methods* (pp. 74-90). New York, NY;: Routledge.

Hockly, N. (2016). Special educational needs and technology in language learning. ELT Journal, 70(3), 332-338. DOI: 10.1093/elt/ccw033.

Keck, C. M., Iberri-Shea, G., Tracy-Ventura, N., & Wa-Mbaleka, S. (2006). Investigating the empirical link between task-based interaction and acquisition: A meta-analysis. In J. M. Norris & L. Ortega (Eds.), *Synthesizing Research on Language Learning and Teaching*.

Keith, T. (2015). Multiple regression and beyond (2nd ed). New York: Routledge.

Lambert, C. (2023). Peronal investment in TBLT. In C. Lambert, S. Aubrey, & G. Bui (Eds.), *The role of the learner in task-based language teaching: Theory and research methods* (pp. 19-40). New York, NY;: Routledge.

Lambert, C., Aubrey, S., & Bui, G. (Eds.). (2023). *The role of the learner in task-based language teaching: Theory and research methods*. New York, NY; Routledge.

Mackey, A. (2012). Input, interaction, and corrective feedback in L2 learning. Oxford University Press.

Merriam, S. B. (1998). *Qualitative research and case study applications in education* (2nd ed. ed.). San Francisco, CA: Jossey-Bass Publishers.

Perinetti, G. (2018). StaTips Part IV: Selection, interpretation and reporting of the intraclass correlation coefficient. South European Journal of Orthodontics and Dentofacial Research (SEJODR), 5(1), 3-5. DOI: 10.5937/sejodr5-17434

Pica, T. (1996). Do second language learners need negotiation? International review of applied linguistics in language teaching, IRAL, 34(1), 1-22. https://doi.org/10.1515/iral.1996.34.1.1

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics, 36*, 50-72. doi:10.1017/S0267190515000094.

Reeve, J., & Lee, W. (2014). Students' classroom engagement produces longitudinal changes in classroom motivation. *Journal of Educational Psychology, 106*(2), 527-540. DOI: 10.1037/a0034934.

Robinson, P. (2015). The Cognition Hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In M. Bygate (Ed.), Domains and directions in the development of TBLT: A decade of plenaries from the international conference (pp. 87-122). John Benjamins. https://doi.org/https://doi.org/10.1075/tblt.8.10sam.

Sato, M., & Loewen, S. (2022). The research–practice dialogue in second language learning and teaching: Past, present, and future. *Modern Language Journal, 106*(3), 509-527. DOI: 10.1111/modl.12791.

Seedhouse, P. (2005). 'Task' as research construct. *Language Learning, 55*(3), 533-570. DOI: 10.1111/j.0023-8333.2005.00314.x.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Taber, K. S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education. Research in Science Education (Australasian Science Education Research Association), 48(6), 1273–1296. DOI: 10.1007/s11165-016-9602-2.

Tarone, E. (2010). Second language acquisition by low-literate learners: An under-studied population. *Language Teaching, 43*(1), 75-83. DOI: 10.1017/S0261444809005734.

Willis, J., & Willis, D. (1996). *Challenge and change in language teaching*. Oxford: Macmillan Heinemann.

7. Please list major findings/deliverables and potential implications of the project that may inform teacher training and/or educational polices related to Chinese/English language learning/teaching.

- Good practices and poor practices when delivering task-based language teaching
- The 5-dimensional TBLT framework for lesson planning and for evaluation
- The actual working with the participating teachers (project meetings and co-planning lesson sessions)
- Proposed definitions of TBLT adapted for SEN schools
- Sharing seminars with workers in the field.