

# 普通話口語測試工作坊問答實錄

吳偉平

## 前言

香港教育署課程發展處在 2002 年底為中小學普通話教師舉辦了一個以普通話教學和測試為主題的講座和三個工作坊。參加教署主辦工作坊的老師提了不少問題，主辦單位認為工作坊上的問答對其他沒有機會參加的老師會有所幫助，要求整理成文。由於當時討論頗為熱烈，所提問題又多，只好分門別類，把一些較有普遍性的問題以問答的形式寫下來，重點是口語測試。雖是實錄，已非原汁原味。

- Q1: 普通話課一定要考口試嗎？
  - Q2: 考試時間不夠怎麼辦？
  - Q3: 考口試時學生不開口怎麼給分？
  - Q4: 甚麼是考普通話聽說最好的方式？
  - Q5: 老師之間評分不一樣怎麼辦？
  - Q6: 開放型題目如何評分？
- 
- Q1: 普通話課一定要考口試嗎？

這是個很簡單的問題，答案可深可淺。簡單回答是「不一定」。真要好好回答可以先從三個反問開始。第一

個：「考試的目的是甚麼？」第二個：「考甚麼？」第三個：「怎麼考？」

有的老師認為不考試就沒有分數，聽起來好像老師的目的就是為了讓學生有一個分數。但是口語考試分數代表了甚麼呢？老師的主觀願望是希望這個分數代表了學生的普通話能力，客觀現實卻未必如此。假如只考看漢字寫拼音，有兩個學生，一個拿了 100 分，另一個只拿了 60 分，我們只能說拿高分的學生有很紮實的漢語拼音知識，不能因此而斷定前者的普通話能力一定比後者強。不具備語言知識的學生當然不會有太好的語言能力，學過的詞彙都唸不好，很難想像這樣的學生能在談話中熟練得體地使用這些詞彙。反過來就不好說了，課文讀得好的人，在生活中無法流利地說話並不少見。也就是說，沒有知識就沒有能力，但有知識卻不一定就有能力。

假如說老師的目的是為了評估學生的普通話能力，考試只是了解學生能力的手段，那麼這個試就不是非考不可了，因為評估學生的口語能力有很多方法（參見 Q4）。

參加工作坊的老師所提供的信息表明，不少中小學普通話課的考試形式目前仍以間接考試為主。簡單的說，就是通過考核學生的普通話知識（辨音、拼音、朗讀課文等）來推斷學生的普通話能力。直接考試（開放性問答題、情景對話等）的比重很低，所以很難在學生的口試成績與普通話能力之間劃等號。既然考試分數不說明問題，考不考試本身也就不是一個重要的問題。

重要的是分清楚普通話知識與普通話能力的區別，從專業的角度了解到底有哪些因素可以作為衡量學生說話能力的指標。如果你認為在課堂上用普通話回答老師問題的能力，用普通話與同學進行討論的能力，根據生活中的實際情形用普通話傳達信息的能力（比如用普通話報告週末做了些甚麼事情）可以作為衡量學生普通話能力的標準並在平時為學生建立檔案，以這些平時分為基礎進行評核，一樣可以達到了解學生普通話口語能力的目的。

## Q2: 考試時間不夠怎麼辦？

比較重要的口語測試，常見的做法是兩個主考官對一個考生，像香港的教師語文能力評核。這種考試形式的長處是主考人員有單獨與考生對話的機會，可以根據考生的表現即場作出評估。最大的短處是可行性（Practicality）低，因為太耗時。中小學普通話上課時間本來就不多，加上老師少，學生多，很難實行二對一的考試形式。就算一個學生只能用三分鐘，有的老師說把所有的上課時間都拿來考試都不夠。在這種情形下只好採用其他方式來評估學生的普通話能力（參見 Q4）。

當老師的要多了解並學會善用不同的方法來評估學生的口語水平，如讓學生參加小組討論，自己在旁邊根據學生的表現給每一個學生打分；讓學生在課餘時間自己根據題目要求錄音，再根據錄音帶上的語料評分等等。專業人員的研究表明，以學生的平時分數為基礎計算出來的成績，往往有比較高的效度。與正式考試的相關係數也相當高。不過這個平時分並不是所有作業分數的總和，而是按照一定的原則為每個學生建立的學習檔案。

老師在採用不同方法的時候，要清楚地看到目的與形式之間的關係。除了具體的考試方法，教普通話的老師對口語測試的一些基本問題也要有所了解，比如考試的效度（validity）、信度（reliability）和可行性，口語考試中考與評分開的可能性，如何通過增加口試的前期製作減少評核時間，如何利用多媒體和高科技等。

總而言之，口語測試有很多不同的方法，每種考試形式都各有利弊。重要的是了解每種形式的理論根據，長處和短處，並根據本校的實際情況採取合適的方式。這樣才能在有需要的時候據理力爭，取得必要的支持，面對來自同事或家長的質疑的時候，還可以講出個所以然來。

### Q3： 考口試時學生不開口怎麼給分？

口試評分的基礎是學生的語言表現，老師聽不到學生滿意的回答會給低分，學生不回答，沒有提供語言素材，當然就有可能拿零分。

問題是，學生在考口試或上口語課的時候不說話或不回答問題原因很多，比如不習慣在全班學生面前說話，性格內向或害羞，聽不懂老師的問題，聽懂問題但沒有回答問題的知識，甚至認為問題不值得回答等等。

口試沒分數或低分，最常見的解釋就是學生的口語水平不行，假如造成學生分數低的原因與上述因素有關，這個分數就不能反映學生真正的普通話水平，用專業的話說，就是沒有信度。

如何讓學生開口說話是一門大學問，也沒有甚麼靈丹妙藥，不過有的老師或許可以從考試形式的問答中（參見Q4）得到一些啟發。

### Q4： 甚麼是考普通話聽說最好的方式？

考試可以簡單分為直接考試和間接考試。就普通話口語測試而言，最直接的考試就是看學生在自然的語言環境中是否能運用普通話進行溝通，間接考試是通過了解學生的普通話知識做出關於普通話能力的判斷。從老師的發言和提供的試卷中，可以發現每個學校都有自己的做法，常見的有：

1. 讀拼音
2. 讀單字，詞語，課文或詩歌（漢字）
3. 選擇題
4. 填空題
5. 語音辨析
6. 聽寫

7. 翻譯，包括粵/普，拼音/漢字，等等
8. 用平時分，不考試
9. 問答題
10. 看圖說話
11. 學生小組活動，老師評分
12. 錄音帶（學生用錄音機錄音，老師通過錄音帶評分）
13. 普通話演講，包括預先給題或即場給題

每種考試形式在具體操作的時候又有很多需要注意的地方。高效度的題型（8 – 13）如何提高信度和可行性，高信度的題型（1 – 7）如何做到有一定的效度，考試形式與教學模式如何配合等等，這些都是主考老師應該考慮的問題。從實用的角度看，考試的方法其實無所謂好壞，不同的方法各有千秋，能達到目的就是好方法。不過老師一定要明白自己所採用的方法是偏重知識還是注重能力，有不少老師其實對知識和能力之間的界線並不是十分清楚。一般來說，開放性問答題最能考出學生的實際說話能力，但評分的難度較大。

Q5: 老師之間評分不一樣怎麼辦？

不同的老師評分不一致是常見的問題，首先得知道不一樣的原因，才能對症下藥，力求一致。填空題和聽寫這一類題型的不一樣很容易解決，一般都是尺度把握的問題，比如用的詞語是否與答案意義相同或相近，學生字寫得太潦草扣分不扣分等，這些都可以通過協商取得一致意見。老師之間各持己見，難以妥協的往往都與開放性問答題有關。

原因之一是老師各有不同的標準，語音、語法、詞彙量、流利程度、答案是否言之有物、所提供的信息正確與否、所用言語與說話人和說話場合是否有衝突，這些都可以成為評分的標準。有的老師重視語音，有的重視語法，對某位考生的評估有爭議，本質上可能是不同的語言觀和

教學觀的衝突，考生的表現只是這種衝突的在具體問題上的體現而已。

就算是語言觀和教學觀相同的老師，對同樣的標準也有不同的解釋和尺度。同是重視語音的老師，有人認為單音重要，每個字的聲母、韻母、聲調都得到位，有人覺得只要能達意，個別音沒到位也可以得分。

參加工作坊的老師有不少人因為同事之間評分不一致而深受困擾，在試圖說服別人之前，最好先弄清楚自己和對方在評分標準和尺度掌握上有甚麼根本的衝突。只要在大的理念上有共同的基礎，同事之間在學生分數上的具體差異應該不難解決。

另一個常見的原因是受平時學生表現影響，先入為主。有的學生平時表現很好，考試的時候馬失前蹄，任課老師給分往往比其他老師的分數高。如果是單獨測試，老師有時比較容易受前後考生影響，中等的學生在好學生後面可能會拿低分，在較差的學生後面則可能會得到較高的分數。這些也可以帶來分數上的不同，是老師要注意的地方。

主考老師之間分數是否一致是測試學中一個必須面對的問題，好的考官之間應該有較高的相關係數，也就是說，一個考生，不管是甲老師還是乙老師評分，其結果應該基本一致，這樣的評核才有信度。問答題是口語測試中效度較高的評估形式，不少老師其實對這一題型的複雜性認識不足，以為隨便出個題目，讓學生說說話就能考出水平。希望關於「開放題目如何評分」的問答（參見 Q6）對這些老師能有所幫助。

#### Q6: 開放型題目如何評分？

開放型（Open questions）題目留給學生自由發揮的空間，是評估學生實際口語能力的高效度題型。老師在評分

時，首先必須就評分標準達成共識，常見的標準可以從三方面討論。第一是答案的內容，學生都說了些甚麼？是否回答了問題？評分時老師之間應該先統一標準。比如描述課室，可以要求學生細講一點，也可以羅列十幾點，不過老師必須都按事先同意的原則評分，不能隨心所慾，才能保證考試的信度。

第二是答案的形式，學生是怎麼說的？重點在語言本身的正確程度，包括語音、語法、詞匯或表現形式（語段結構、言語組織、邏輯、條理等）。有的老師喜歡把語言形式量化，規定錯了幾個音就扣多少分，好像規定得越細就越科學。其實開放題目的重點是內容，有了太多關於語言形式的具體規定，評分時往往拘泥局部而忘了整體，反而把內容放到次要的位置上，這是本末倒置。

第三方面是學生所給的答案是否得體，考的是學生量體裁衣的能力。語言水平高的人，不管用普通話、廣東話還是英文，都會注意語境問題，包括說話對象（師長、同學、朋友、弟妹等），說話的場合（正式、非正式等），恰如其分地使用成語、俚語、禮貌用語。

不管是甚麼水平的考生，老師都可以按照第一、二方面的標準進行評估，不過往往只有高水平的考生，才能有足夠的語料讓老師從第三方面進行評估。儘管如此，老師也要注意從低年級開始培養學生對語境的敏感性，因為語言的運用始終離不開特定的語言環境和上下文。既然把語境作為評核的標準，出題時就應該考慮如何合理地包括語境因素，這已經不是三言兩語可以說清楚的問題了。

## 作者簡介

吳偉平博士，香港中文大學新雅中國語文研習所所長，曾任美國應用語言學中心漢語測試專案主任，以語言學中心研究員身份參加美國教育部資助的大型語言水平測試研究專案（1992 – 1999），負責一系列漢語能力測試及培訓教材的設計及修訂工作，並為美國各大學和中學漢語教師舉辦專題培訓班（語言水平為綱的教學與測試）。吳博士也曾應邀為聯合國設計漢語教師入職資格試並任主考，為美國國家法院中心設計法庭翻譯資格測試（普通話、廣東話）並任主考。

吳博士在香港為考試局等機構提供諮詢服務，並應邀為教署、國際學校等機構舉辦不同形式的研討會和工作坊，並為不同群體設計普通話水平測試，其中由研習所研製開發的 Q 版普通話水平測試是語常會與香港推廣普通話大聯盟 2002 年普通話月的一個主要項目，曾在全港 18 區舉行。