

# 台湾学生成就评量资料库之建置与展望

台湾师范大学教育心理与辅导学系

国家教育研究院测验及评量研究中心

李俊仁

## 一、TASA 的主要业务

今天主要内容是集中于 TASA 的主要业务，以及 TASA 有助于各个地区的学力检测，台湾关于国民基本教育所面临的基本问题：台湾现在一个很重要的转变是从九年的义务教育转成 12 年的国民基本教育，第一个很重要的原因是学生的竞争力，台湾面临开放市场的国际竞争的情形，现在人力资源上最大的竞争对手实际上是大陆，这是我们不可避免的事情。

第二是台湾少子女化以及老年化，每个学生的产值必须比过去高。现在台湾 65 岁以上人口占总比是 10.7%，到退休的时候（2030 年）是 24.4%，现在的台湾人口平均寿命是男生 72 岁，女生 83 岁。2060 年是 41.6%。现在一个人负担两个，20 年后是四个，再来 40 年是 6 个。现在教的学生决定了我们将来是否拿到退休金，这是非常严重的的一个讯息。台湾还有其他的教育状况，是做教育研究的人回去思考的问题。不仅是考虑国际的评比，学生的平均数字，我们想看高分和低分的差距，统计学上的标准差。媒体记者可能只关心排名，做教育的人关注最弱势的学生，他的学习状况是怎样的情形。特别在阅读表现里，我们和世界各国去比呢，高能力等级的学生的百分比相对上是少的，在低能力百分比上是相对多的。虽然台湾和国际的平均值比很好，但强有力经济竞争国、地区是香港、新加坡和中国大陆、韩国、日本，所以我们不但要和 OECD 平均值比，要和上述国家和地区比，才有竞争力。在国际 PIRLS 的比较中，有分不同的等级，最低等级是 400 分。400 分的概念是指，如果在文章里写的清清楚楚的讯息，我们的学生有没有办法取得？“在这张粉红色的桌子上有一张白开水，请问一下在粉色的桌子上是什么东西”，就这么简单的问题，有 2% 的学生没有办法达成，是无法达到最低基准的。这是针对的四年级的小朋友。下面针对的是 15 岁的学生，我们的国三跟高一或专一的学生，这些资料公布中，我们还是关注在国际的排名中的上升抑或是下降，还是一样，我们做教育的还是要知道做不同类比的情况，在 PISA2009 的调查中，在做调查之前，通告了一个所谓的测验架构，这个架构中分成了学生不同的能力等级，数字越低代表越不好，数字越高代表越好的情况，一样呼应了 PIRLS2011 的调查，算是比较早的调查，我们学生将来有没有应付咨询阅读的基

本能力？有 15.4% 的学生是有一定困难的，有很清晰的数据告诉我们这些困难点的状况，我们还有关于地区差异的数据部分，这不是我个人的资料，是从研究员身上的资料把他统整的结果。我们阅读的时候会需要基本的识字能力，发展一个简单的指标去测量一个小朋友的最低识字情况，然后推估出来他基本阅读时需要的最低的识字量。图表显示的是台湾一到九年级的学生识字的状况。可以明显看到全台湾和花东地区的差距，假设我们把文章里 99% 字的覆盖率是最基本的识字量，独立阅读最低基本识字量为 2708 个，在花东地区 9 年级大概 8% 的学生在阅读基本识字量上是有问题的。

另外一个资料是国中会考的资料，国中会考是台湾 9 年级要考的考试，在这个国中会考的时候有所谓的标准设立，有“精熟、基础和待加强”三个等级。在各个科目考试中“待加强”的比例基本 15%-20% 左右，这个比例不低，假设“基础”真的是我们设定的最基础的等级，代表我们有接近五分之一的学生在这类科目上有一定困难。我们怎样去解决，也许有人会说我们设的“基础”太高了，这也是台湾经常发生的状况，但个人如果在现场一定会有这种感觉，特别是最近在和现场一些老师开会的时候，他们说我们定的一些基准，连他们都做不到。我们课纲里面可能塞了太多的东西，标准设得太高，特别是偏乡的一些小朋友根本没有办法达成这样的基准。但不管怎样，我们是有一部分的学生无法达到最低基准，这是事实。所以我们希望不是就评量而评量，评量当然只是一个手段，我们希望帮助每一个学生，像《NO CHILD LEFT BEHIND》报告一样，每个学生都达到一定的基础的情况。在各级教育中，从幼稚园开始，我们有没有办法让待加强的比例降低，达到拔尖、扶弱，全面提升的作用。

所以我们有一些基本要求要去完成：一是标准的设定，也就是每一个年级，需要达到的最低基准，至于这个最低基准是什么，功能都可以讨论，但必须设定一个最低基准，二是基准需要评量工具与机制，我印象中上海教科院是要发展评量的工具，三过去的评量很大基础上是 assessment of learning，学习的成果是什么样子，而现在是 assessment for learning，帮助小朋友，他的困难点在什么地方，应该有什么样的教学活动，设计什么样的教材，老师应该提供什么样的方法。是在评量的过程中让学生去学，各位应该听说过有“翻转教师”，基本概念是老师把上课的基本内容先播放给学生看，学生在家先去看录影片，学生在学校基本上是针对他不会的东西来和老师讨论，老师可以布置一定的作业。也就是说学习在家里，评量在课堂上，这是 20 年之后的趋势，教师的专业智能：知

道学生的困难点，给予指导和帮助，能不能够提供足够的教材与教导的方式，这绝对是 20 年后教材发展的方向和趋势，这是一个重要的改变，四怎样去保证每个学生都能发生这种改变，就是成长监控，不同的时间段落里是否有办法达到最低的成就指标？如果没有达到的话，就要回到（上述四个基本要求）里面做一个循环运作的过程，坦白说，这并不是我提的新的概念，而是应该要去做的事情。

两种评量的内容，一个是学科成就的评量，一个是能力指标的评量，过去我们比较着重的是学科成就的评量，就是老师教什么，我就考什么，所以学生只要针对教材精熟就好，通常考试的内容是能否填充课文的内容，能否默书等。除了成就的评量之外，也需要能力的评量。在台湾的人可能会有这么一个经验，“天这么黑，风这么大，爸爸捕鱼去，为什么还不回家。”大家应该都还会背。这是各位可能当年所熟悉的课文，我到现在还会背，可见当年花了多少时间去背，这是课文能力的成就上的一个评量，有没有办法在应用上抽象化或综合分析的能力，在学科的评量上是国语的评量，能力上的评量可能是阅读，以阅读（能力非学科）而言，因为我自己是做阅读的，要发展这些基本的核心的监测方式，包括识字量，每分钟朗读字数，有没有足够的词汇，可以测识词量，有没有足够的阅读理解策略。每一个都对对应怎样学或怎样教的内容。

举一个例子，有一个识字量评量，假设今天遇到一个五年级的学生，他的识字量 1200 字，大概二年级小朋友的平均识字量。你要怎样教他，是教二年级的教材内容吗，怎样都说不过去，学五年级教材的内容对他来说又太难，第一是怎样挑到合适他的教材去学，第二是假设目标是增加他的识字量和识词量的时候，需要教什么样的内容，有一个东西叫做基本指代式教学法，心理语言学里的一个概念叫做组织规则，能否通过所有部首的字教学，比如所有偏旁是青色的青字，合在一起教他，并且不仅是教学，而是把所有词和句子一起带进来。

评量的目的不是为了评量，目的是帮助每个学生学习。

## 二、台湾 TASA 评量的内容。

接下来和各位介绍一下台湾两个比较大的评量的内容。一个是台湾学生学习成就评量资料库（TASA），另一个是县市举办的学力检测。刚才已经说过，我们在评量的时候，要区分成就的评量和能力的评量。TASA 设置的目标最主要的就是两个目标：

（一）九年一贯课程纲要订定能力指标的学生实际表现状况。起到一个监控的情形，当然除了监控之外，还要提供反馈，反映出来学生实际的状况。

(二) 建立学生学习成就长期资料库, 以检视课程与教学实施成效, 提供政策参考。在台湾有一件事很多人都凭感觉说, 现在学生的能力越来越糟, 这有没有实际上的资料。假设你讲的是大学生的素质的话, 我那个年代大学生只有两万人, 现在是十万了。如果今天去台大电机系, 说他们的素质下降的话, 我会承认你说的下降。所以你必须考虑一些很基本的指标。同样的, 我们马上要推 12 年国教, 很多人会担心我们把考试放掉了, 学生的能力会不会下降? 一样国家要有一个指标告诉我们学生能力是不是有下降。TASA 可以提供这样的指标。所以实测的东西可以提供教育思考的情况。主要是小四、八年级和十二年级的学生, 在台湾九年级的学生是不太可能来考的情况, 必须需要往下移。所以我们集中在小四、小六, 国中二年级跟高二的情况。在测的科目方面, 2005 年由五个大学负责不同科目, 那时产生了一些的问题。从 2008 年开始, 由国教院承办。自 2009 年开始, 国教院开始正式的检测, 那时是小学四年级、六年级第一轮, 隔一年就是国二, 再隔一年就是高二, 每三年一个年段的调查, 去追踪他的成长变化的情形。在实施的时候, 每个学生是参与两科的设置。然后需要做背景问卷, 这几年又多了校长跟老师的问卷帮我们了解运作的一些状况。这个评量和一般的评量有点不太一样的情形是反映课纲所订的内容, 学生的学习状况会怎么样, 会提供 100-200 题, 但每个学生坦白讲只能考 30-40 题, 所以会做一些调整。我们在检测完成之后, 大致上会有三项基本作为, 我们会写三个报告, 第一个是评量结果报告: 次年向教育部部次长报告结果; 第二个是试题分析报告: 与中央以及各县市辅导团分享教学建议, 提供指定课纲时的实证参考资料。实测的内容和学生大致上的状况。最后一个技术手册报告: 如何抽样和加权, 结果是怎样。公开原始资料, 任何研究人员有兴趣可以通过申请使用资料进行研究。

可提供的讯息: 刚才讲到开始是每年做一次, 后来是每三年做一次的状况, 以下这是示意图, 如果要显示确切数据, 必须打公文, 被批准后才可以使用。地区的差异是指台湾东、中、北部的差距, 有一些基本的问题可以问, 第一个是整体学生能力变化, 每一个学生在每一科里历年变化的情形。第二个是地区别, 这是个很基本的大家会去关心的问题, 第三个不是那么重要但大家一定会关心的就是性别差异的问题。还有在台湾比较特殊的是身份别(原住民、新住民子女)的考量, 他们的成绩和一般学生成绩的差异, 实际上社经地位和一般民众子女是有一定差距的。还有都市化程度, 台湾各区有不一样的都市化程度, 高度发展区、城镇和偏乡, 这三组在年度上的差异是怎样的。做教育的都会问一个很基本的问题

题就是社会经济地位，不幸的是我们不太能够直接问，一方面小学四六年级的学生他也不知道，另外一方面，实际上有一点点敏感，也不太能够直接问的情形。这个跟美国一样，美国也不能直接问他们的社经地位，只能问有没有 a free lunch, 代表社会补助，通过有没有免费午餐的方式推测，只能以推估的方式，无法直接调查；如家里有几本书等讯息，间接地推论出社经地位状况。

这里面重点的讯息是：我们有一般生和特殊生的情况，差距可能必然存在，都市和乡村一定有差距，这是不可否认的事实，我们考虑的时候是政策到底做得怎么样，但差距是否跨大、持平还是缩减？台湾新住民的子女，和一般子女的差距是在缩小的，几乎快要没有，这是我们很高兴看到的状况。

这些可以提供的资料，但具有困难点：可以提供描述性资料，但无法提供解释；描述性的资料：中譬如说学生的能力有没有下降，是持平，或是缓慢提升；有一些解释性资料，相对我们是没办法回答的，-造成原民学童成就低于非原民的原因，我们也知道他和一般生有一定的差距，问题是怎样缩小这样的差距，很难提供一些解释性的资料，但是这并不是 TASA 特有的问题，是全球所有做大型资料库都面临这样的一个问题。

TASA 未来展望面临一定的压力：从 9 年义务教育转成 12 年国教，当年评估课纲的部分，要做一定程度的变化，一些代表希望 TASA 能够跟补救教育结合。未来的发展是希望有更清楚以及合理的标准设定，设定各个年级的最低成就标准；希望能走向素养调查，而不是原来的成就调查，另外一个是因应国际的趋势，走向数位评量；数位阅读很重要的是怎样去做搜寻，怎样找到关键词，搜寻之后怎样排开无关的学习：希望能够跟国际评比架接，比如 PIRLS。

### 三、TASA 协助县市进行学力检测

目的，各县市不同：有的是想提供教学辅导建议，有的是筛检出低成就者。有的是想了解区域差异。

“教育部”有颁布一个检测的要点，能力检测结果不得以个别学校、班级之形式公布，最多拿到全县和自己学校的结果，没有办法知道其他，也不得作为教育人员绩效考核、奖惩之依据，校务评鉴之项目，这方面香港也完全一样。

做学力检测的时候有两种方式，一种是县市自办，另一种 2014 年由 14 个县市由国教院协助，大概是学生人数的三分之一或者四分之一。主要是四年级和五年级，科目是国语文和或数学，部分县市有其他科目和其他年段参与。

TASA 协调各个县市，训练一些老师能帮我们出题。我们再做组题和出题的

工作。施测行政由各县市负责，国教院是出题和提供现场资料，实测后由国教院分析资料，提供分析报告，包括各县市的通过率，每一年的鉴别率、教学的建议。所有试题为抛弃式，并没有进行年度等化。

#### 四、补教教学评量展望

台湾现在有一个非常重要的计划，叫做补救教学，过去花了台币 6、亿，这两年大概花了 12-15 亿来进行。早期并没有评估他的绩效，后来开始委托台南大学开始去做，国家为了帮忙那些学习成就低或者趋于弱势的小朋友，我们设定一定的基准，假设他在颁布区域里面 PR 小于 35 进行数位评量（但低年级否），依是否通过最低基准，决定是否参加补教教学。一到九年级，语数英三科。每年九月基本测试，二月以及六月，追踪测试。由课后补教教学，到最近“教育部”提出一个方案是严重落后学生是课中补救教学。比较困难的是教师资源

在展望部分，怎样通过测验知道学生的能力，学哪些是教材，用什么样的学习方式，老师怎样帮助这些小朋友。最后我们在谈补救教学时，要制定这些指标，一定要到现场去测看学生的情况，根据测的情况再回去修指标，是一个轮回的过程，希望知道学生到底在这样的题目下通过率是一个怎样的情形，让更多的学生通过教学评价的方式学习。

[Jrlee@mail.naer.edu.tw](mailto:Jrlee@mail.naer.edu.tw)