

新加坡中小学华文课文的语文特质与可读性初探

苏启祯

摘要

语文课文的语文特质与可读性是语文教学成功的基本条件。英文可读性公式甚多，且常用于检验英文课文的难度，而华文可读性公式较少，并且主要适用于华文为第一语文的文本。本文首先分析新加坡小学华文课文的语文特质，然后采用新近开发的华文二语可读性公式检验了不同年级的课文的难易度。之后，笔者选取了中学三种华文课程的课文，比较了不同课程以及不同年级的语文特质与可读性。结果显示，中小学课文的可读性受字词相关的特质影响较大，而受语句相关的特质影响较小。此外，小学课文的可读性按年级递增，但中学课文缺乏规律。

关键词：语文特质 可读性公式 华文作为第二语文 回归分析

语文教学的主要工具是语文课本。因此，检查与控制课文的语文特质（Language features）和可读性（Readability），力求配合预设的年级，是语文教材编纂和运用的基本原则。如果课文的语文特质和可读性超过预设年级学生的能力，教师和学生将面对诸多困难，对学习动机和成绩会有不良的影响。一般来说，在缺乏科学化工具检查语文特质和可读性时，课文撰写人员只能凭个人的经验去判断课文是否适合预设的年级。然而，个人经验因人而异，其中主观成分无法控制。同理，教师选择阅读材料时，如果缺乏科学化的工具检查语文特质和可读性，便会面对同样的问题。

课文复杂性（Text complexity）必须兼顾文本的质和量。美国《共同核心州际标准》（*Common Core State Standards Initiative*, 2019）提出文本复杂性（Text complexity）包括三个范畴：(1) 文本的质，如语义、结构、惯例、明晰、知识要求；(2) 文本的量，如可读性及其他复杂性数据；和 (3) 读者与文本及任务的配搭。本文将从课文的质和量两个方面分析新加坡中小学华文教材，即运用量化分析方法展示不同年级以及不同课程教材的语文特质和可读性。

语文特质

新加坡华文教研中心的《华文教学资源平台》（林进展、赵春生、洪瑞春、吴福焕、王志豪，2015）是本研究用以分析语文特质的工具。该平台有多种功能，包括单字查询、词语查询、句型查询、文本分级、全文查询、教师论坛等。本文运用该平台的全文查询功能，运算以下八项语文特质：

1. 总字数：课文单字的数目，即篇幅的长短
2. 字种数：课文中不同字的总数（重复出现的单字不计）
3. 总词数：课文中词语的总数
4. 词种数：课文中不同词语的总数（重复出现的词语不计）
5. 整句数：以句号为根据计算的句数
6. 小句数：以逗号为根据计算的句数

7. 最大句长：语句最长的单字数
8. 平均句长：所有句子字数的平均数

以上八项语文特质，前四项和字、词有关，而后四项则和语句有关。可预料到的，语文特质之间会有相关。文本的总字数多，它的字种数、总词数、和词种数也会跟着多。同理，整句数大，小句数、最大句长、和平均句长也就随着大。就年级而言，较高年级的课文，各语文特质的量也可能较大，课文的难度也随之提高。同时，值得注意的是，字和词是表达课文内容的语文特质，由于课文的话题（如日常生活、科学现象、或历史人物等）有很大的不同，因此使用的数量和遣词用字方面，差异也大。另一方面，句子类型的差异比字词的差异少得多，变化也就不大。从统计数据的观点分析，差异大的变项会有较大的标准差，因此，在回归分析中的预测能力也比较高。据此，本文假设和字词有关的语文特质对课文可读性会有较大的影响。

可读性

运用可读性公式（Readability formula）预估文本难度或适用年级，在英文非常普遍。从最早 1948 年的 Flesch Reading Ease 到 2006 年的 Strain Index，有不少于 23 个不同的阅读公式（Bravos, 2010）。这些公式，基本上是将语文特质量化，通过多元回归分析（Multiple regression），选择较能预测阅读难度的语文特质，加以保留，建立可读性公式。例如 Flesch Reading Ease 采用的是平均句长和每字平均音节数，Coleman-Liau Index 采用的是每 100 字的平均字母数和平均句长，而 SMOG Readability Formula 采用的是多音节字（三个音节或以上）的字数。一般来说，不同可读性公式评估的结果是比较接近的，例如，笔者将安徒生的《国王的新衣》输入七个不同公式，大多数结果显示该篇文章比较适合十二、十三岁（六、七年级）的儿童。

英文可读性公式的应用，有相当长的历史，然而，华文可读性公式相对缺乏。最早的是 Yang (1971) 在美国 Wisconsin 大学设计的，后来有孙汗银（1992）设计的。王蔷（2008，页 51）综合 1971 年至

2005 年的研究结果表明，能够预测汉语阅读材料可读性的语文特质有五项：字均笔画数、词表 / 词频、词数、句均字数、和句字数。但是，王蕾（2008，页 50）也指出，在汉语作为第二语言教学领域中尚未出现客观的量化的可读性公式。最近，宋曜廷等人（2013）比较了回归分析和支向量机 (Support vector machine) 分析方法的相对效力。逐步回归分析 (Multiple regression) 的结果显示，难词数、单句数比例、实词频数、和人称代名词是重要的预测变项。支向量机分析的结果表示，重要的预测变项有难词数、二字词数、字数、和笔画字中数等。

以上提到的华文可读性公式，都是基于华文为第一语文的文本。新加坡的语言环境十分复杂，目前以英语为主流语言，华文为第二语言，华文教材也都是针对第二语言学习者编写的。此外，由于新加坡的华文有许多地方性的词汇和表达方式（语句）。因此，现有的公式可能不适用于新加坡。笔者采用新加坡华文教研中心的《教学资源平台》，开发了针对华文作为第二语言文本的可读性公式。本研究以小学一至六年级的课文为轨范 (Criterion)，用上述八个语文特质为可读性的预测变项 (Predictor)。本文运用 60 篇新加坡小学华文课文（每级 10 篇），作为开发公式的数据。其中 42 篇为设计公式用，余下 18 篇为检验公式的预测功能。经过四轮的回归分析，最后所得的可读性公式如下：

$$\text{可读性} = -0.163 + 0.011 * \text{总字数} + 0.022 * \text{字种数} - 0.026 * \text{词种数} + \mathbf{0.074 * 平均句长}$$

上列公式显示，八项语文特质中，只有四项有预测可读性的功能，即总字数、字种数、词种数、和平均句长。这表示，总字数、字种数和平均句长量越大，课文难度也越高，比较适合较高年级。然而，词种数越大，课文难度反而降低，这意味着单字如果构词能力强，课文也就比较容易读懂。

需要注意的是，上列公式是计算轨范（文本适用年级）用的，其中各系数的高低并不表示特质影响力大小。例如，总字数的系数（0.011）只有字种数的系数（0.022）的一半，造成总字数系数（0.011）的影响力为字种数的一半的错误印象。此外，平均句长系数（0.074）为总字数系数的七倍，也会造成前者影响力特别大的错误印象。这是因为上列公式中的系数并未标准化。要比较不同语文特质系数的影响力大小，必须用标准化系数。下列是用标准化系数的可读性公式：

$$\text{标准化可读性} = 0.911 * \text{总字数} + 0.744 * \text{字种数} - 0.792 * \text{词种数} + 0.141 * \text{平均句长}$$

据此，特质的标准化系数显示总字数影响力最大，即篇幅越长，课文越难。相对而言，语文特质对课文难度的影响，依序是总字数 > 字种数 > 平均句长，而词种数系数为负数，影响力和字种数很接近，但有抵消难度的作用。基于以上有关语文特质和可读性的理解，这项研究将采用科学化的方法，客观验证新加坡主要华文课程课文的水平。

研究方法

课文样本

近年来，为适应学生不同的程度，华文课程多样化，课本也随之不同。这项研究所分析的课文，随机取自不同课程的现行课文，如表1所示。小学有六个年级，中学四个年级。中学方面，主流的课程是快捷华文课程。水平较高的学生会修读高级华文课程。中学华文特别课程（前称华文B，Chinese Language B）是为母语为非华语的学生而设，允许在小学阶段没有修读华文的学生选读。由于选读本课程的学生多为华文初学者，因此本课程是从外语教学的角度拟定学习目标与教学要求的（新加坡教育部，2015），适合水平较低的学生。小学华文课程取样比较大，是为了通过课文特质的分析，建立华文二语阅读

公式，用以预测个别课文的可读性（阅读难度）。其他课程取样较小，但由于随机取样，所选取的课文应该能代表有关年级课文的特质与可读性。

表 1：课文样本

课程	每年级篇数	总篇数
小学华文	10	60
中学快捷华文（主流）	4	16
中学特别华文（较低）	4	16
中学高级华文（较高）	4	16

分析方法

课文选出后，首先用《华文教学资源平台》分析其语文特质。然后，根据所建立的可读性公式，验证个别课文的难度。由于本项目所关注的不是个别课文的语文特质和可读性，而是各课程各年级的相对关系，以所选用的课文样本为代表，用平均数和标准差进行描述，并通过效果强度确定差异的大小。具体的研究问题如下：

1. 中小学各年级课文的语文特质有怎样的变化趋势？
2. 如果以主流课程（如中学的快捷课程）的课文为基准，其他课程的语文特质是否能配合课程目标？

分析结果

小学华文

表 2 展示了新加坡小学华文课文的语文特质和可读性。如表所示，年级之间的差异明显地出现在和字词有关的特质方面（总字数、字种数、总词数、词种数），而和句子有关的语文特质（小句数、最

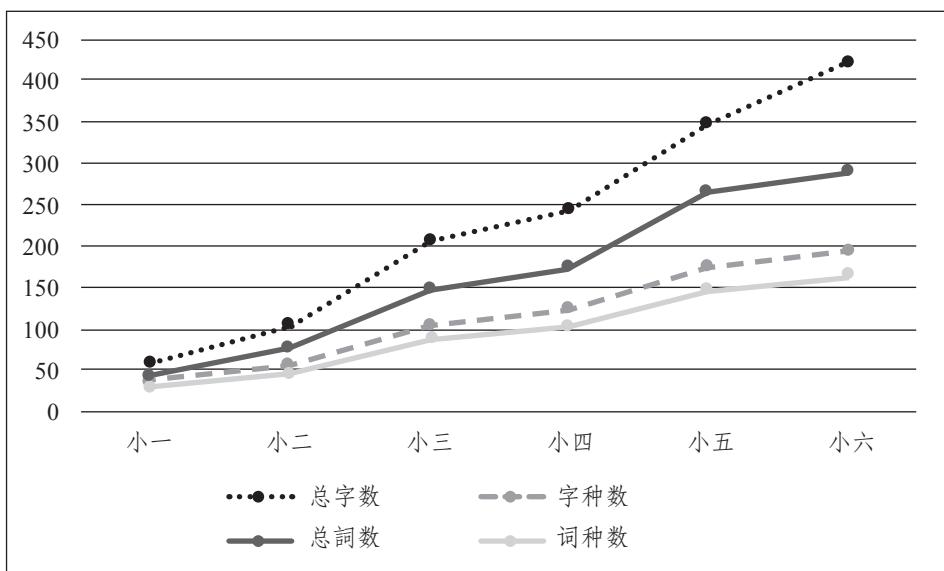
大句长、平均句长），差异很小。这意味着课文的难度受字和词的影响远高于句的影响。具体而言，有关字词特质的差异，最小差异的是 134（词种数），最大的是 362（字种数），而有关句子的特质，最小的差异是 9（平均句长），最大的是 44（小句数）。

表 2：新加坡小学课文特质与可读性

年级	总字数	字种数	总词数	词种数	整句数	小句数	最大句长	平均句长	可读性
小一	59	36	43	30	5	10	18	12	1.3 (1)
小二	104	56	77	48	8	15	21	14	2.0 (2)
小三	207	103	149	88	12	29	34	18	3.4 (3)
小四	243	123	173	103	13	33	38	19	3.9 (4)
小五	349	175	267	146	16	44	46	22	5.4 (5)
小六	421	194	290	164	21	54	45	21	6.0 (6)

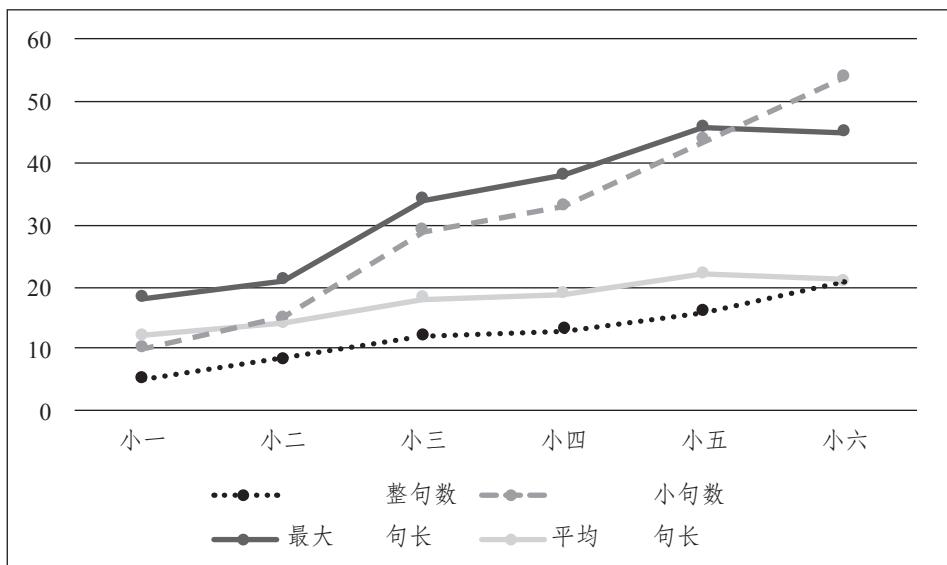
如图 1 所示，从小一到小六，和字词有关的语文特质逐年增加，表示课文的难度逐年提高。但是值得注意的是，增加的幅度并不一致。在总字数和总词数方面，从线条的倾斜度可以看出，在小三之后，有比较急速的增加。

图 1. 小学各年级课文与字词有关的语文特质



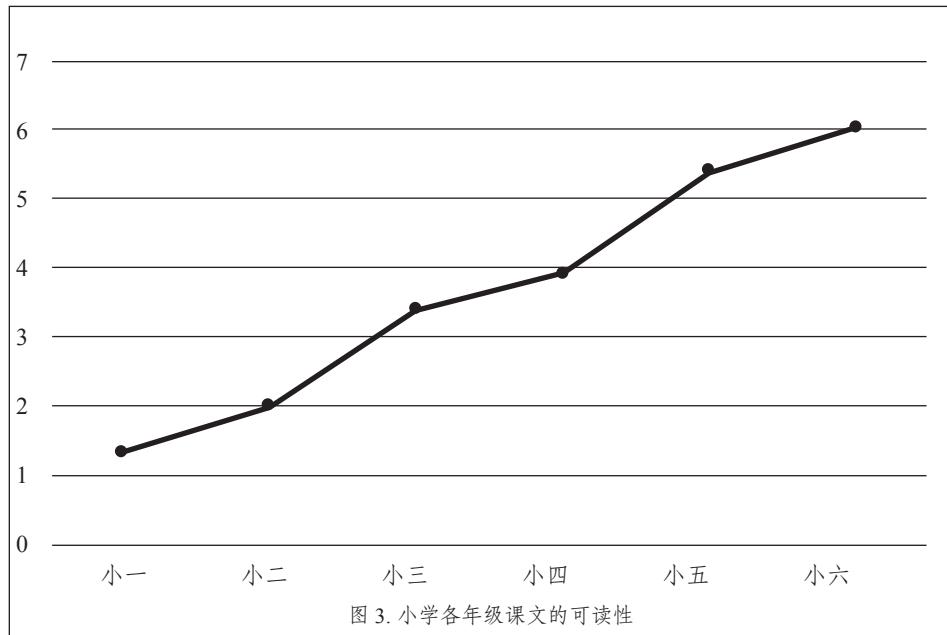
如图 2 所示，最大句长和小句数，从小二开始，有急速增加的趋势，整句数和平均句长的增加则比较平稳。值得注意的是，从小五到小六，最大句长和平均句长的数量有稍微减少的趋势。

图 2. 小学各年级课文与句子有关的语文特质



在图 3 中，Y- 坐标的数字表示每一学年的进度，并假设每一年级的课文难度平均分配于 12 个月。可读性方面，各年级的指数显示课文符合各年级的水平，例如一年级的可读性指数为 1.3，表示课文的平均难度在一年级第四个月，而二年级的课文难度则在该年级开始之处，余者类推。但是，四年级的课文难度 3.9，略低于预期的难度。如图 3 所示，小二到小三和小四到小五的可读性的增幅比较大，意味小三和小五的课文可能比较难。

图 3. 小学各年级课文的可读性



中学快捷班华文

新加坡分流制度下，中学快捷源流学生占中学生总数约 80%，代表主流课程。如表 3 及图 4 所示，中学各年级之间，与字词有关的语文特质差异较大，而与句子有关的则较小。具体而言，中一至中四，与字词有关的语文特质差异在 43（词种数）至 118（总字数）之间，而与句子有关的语文特质差异在 0（整句数）至 13（小句数）之间。

表 3：中学快捷课文的语文特质与可读性

年级	总字数	字种数	总词数	词种数	整句数	小句数	最大句长	平均句长	可读性
中一	601	253	404	224	29	73	42	22	7.8 (8)
中二	632	281	427	246	29	73	52	23	8.3 (8)
中三	719	304	478	267	29	86	53	25	9.4 (9)
中四	716	291	467	258	29	79	49	26	9.3 (9)

如图 4 所示，有关字词的特质，逐年增加，但增幅不大。但是从中二到中三，有比较明显的增加。从中一到中四，字总数和字种数增加略多于 100。如图 5 所示，和句子有关的特质，小句数略有起伏，其余却很少增加。

图 4. 快捷课文与字词有关的语文特质差异

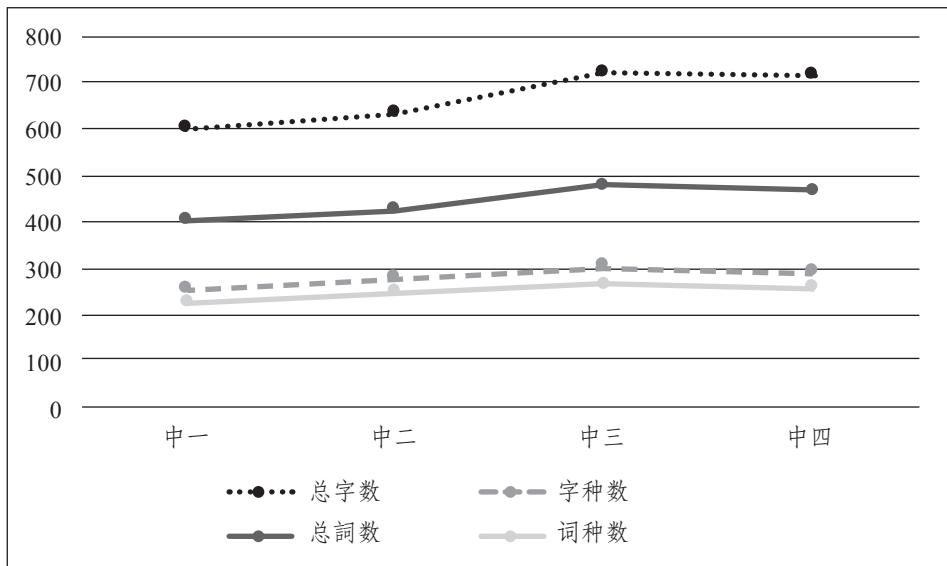
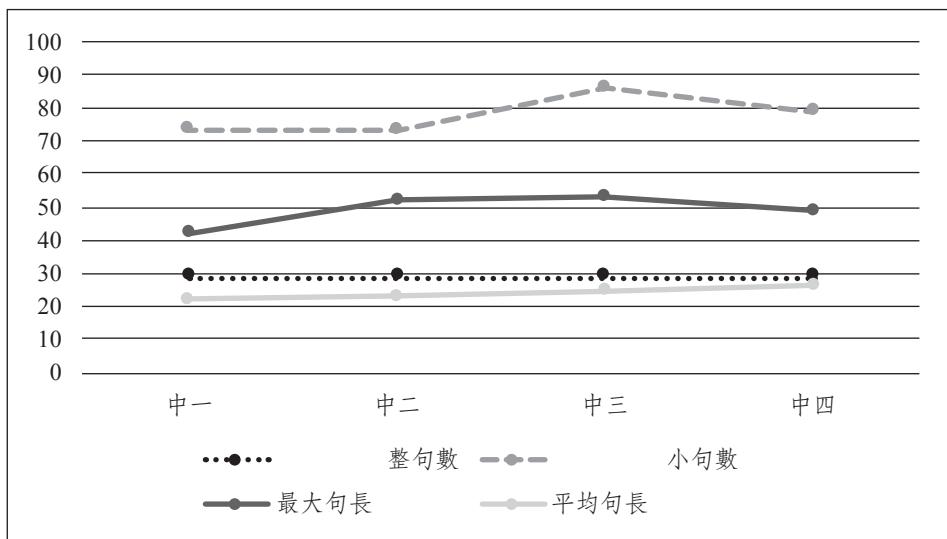


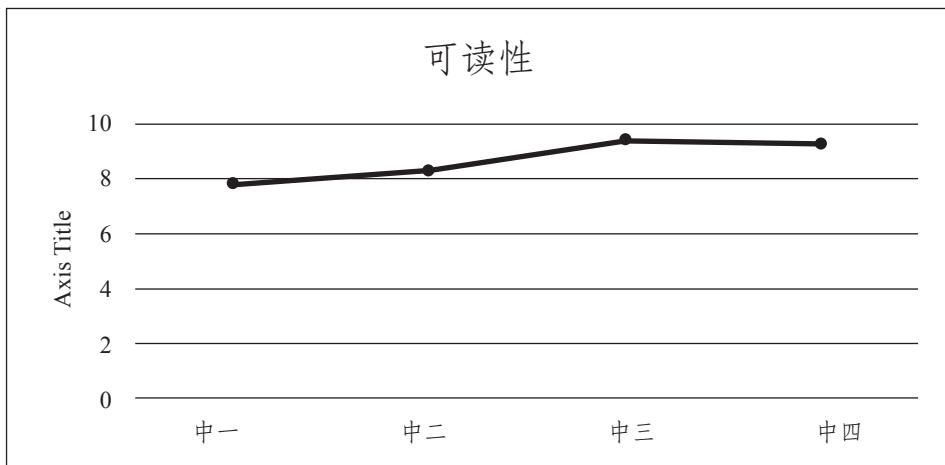
图 5. 快捷课文与句子有关的语文特质差异



在图 6 中, Y- 坐标的数字表示每两学年的进度, 并假设每两年级的课文难度平均分配于 24 个月。从中一到中四, 差距应该有四年, 但是, 如图所示, 实际上只有略多于一年的差距, 从 8 年级到略高于

9年级。值得注意的是，中三和中四的可读性相同，都只略高于9年级，没有差别。

图6. 中学快捷课程课文的可读性



中学特别课程

中学华文特别课程适合水平较低的学生，因此特别课程的华文水平理应明显低于快捷华文的水平。如表4所示，在中学快捷课文和特别华文课程中，各年级之间，与字词有关的语文特质差异较大，而与句子有关的差异则小。具体而言，中一至中四，与字词有关的语文特质差异在42（词种数）至118（总字数）之间，而与句子有关的语文特质差异在4（平均句长）至13（小句数）之间。

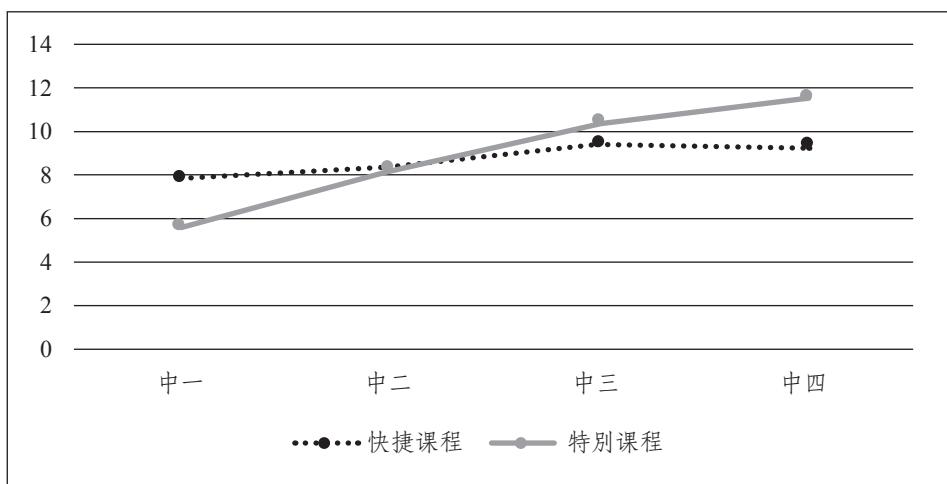
具体而言，两种课程的中一课文有合理的差异，尤其在和字词有关的语文特质方面，效果强度都很大（系数介于1.4至4.4），而和句子有关的语文特质也有些差异，尤其是句长方面。然而，中二课文之间没有明显的差别，除字种数（114）、词种数（110）和平均句长（3.4）之外，其余五个特质差别不大，相应的效果强度也小。此外，两种课程的中三和中四课文特质出现不少负的差数，显示特别课程的课文比快捷课程的难，尤其是字种数、词种数，和整句数。这样的负差会影响特别华文课程中的中三和中四课文的可读性。

表 4：快捷课程与特别课程课文语文特质与可读性比较

	总字数	字种数	总词数	词种数	整句数	小句数	最大句长	平均句长	可读性
快捷中一	601	253	403.6	224.3	29.3	73	41.7	21.6	7.8
B 华中一	436.3	139.3	306.3	116.5	30.8	65.3	36.8	13.9	5.7
差数	164.8	113.8	97.3	107.8	-1.5	7.8	5	7.8	2.1
效果强度	2.2	4.4	1.4	4	-0.2	0.4	1	1.8	3.5
快捷中二	632	281.3	427.1	246.1	28.6	73.1	52.3	23	8.3
B 华中二	617	167.3	413.5	136.5	31	71.8	48.5	19.6	8.2
差数	15	114	13.6	109.6	-2.4	1.4	3.8	3.4	0.1
效果强度	0.1	4.3	0.1	3.1	-0.3	0.1	0.4	0.9	0.1
快捷中三	719.4	304.3	478.3	266.7	28.7	86.1	52.6	25.3	9.4
B 华中三	770.5	186.5	514	153.3	33	82	41.3	24.7	10.3
差数	-51.1	117.8	-35.7	113.5	-4.3	4.1	11.3	0.6	-0.9
效果强度	-0.5	3.3	-0.4	2.9	-0.8	0.3	1.5	0.2	-1
快捷中四	716	291	467	258.3	28.9	78.9	49.1	25.8	9.3
B 华中四	890	190.8	607	164	36.5	93.5	49.8	26.2	11.5
差数	-174	100	-140	94.3	-7.6	-14.6	-0.6	-0.5	-2.2
效果强度	-1.6	2.3	-1.7	2	-1	-1	-0.1	-0.1	1.8

在可读性方面，如表 4 和图 7 所示，特别课程中一课文的可读性指数比快捷课程的低了 2.1，表示两课程的难度相差两年。这符合特别课程设计的初衷，课文适合华文水平较弱的学生。但两种课程的中二课文难度差异很小，只 0.1 年（约一个月）的差别。然而，中三和中四的课文，特别课程的反而比快捷课程的高出两年。这显然不符合特别课程设计的目标。

图 7. 快捷课程与特别课程课文可读性差异



中学高级华文

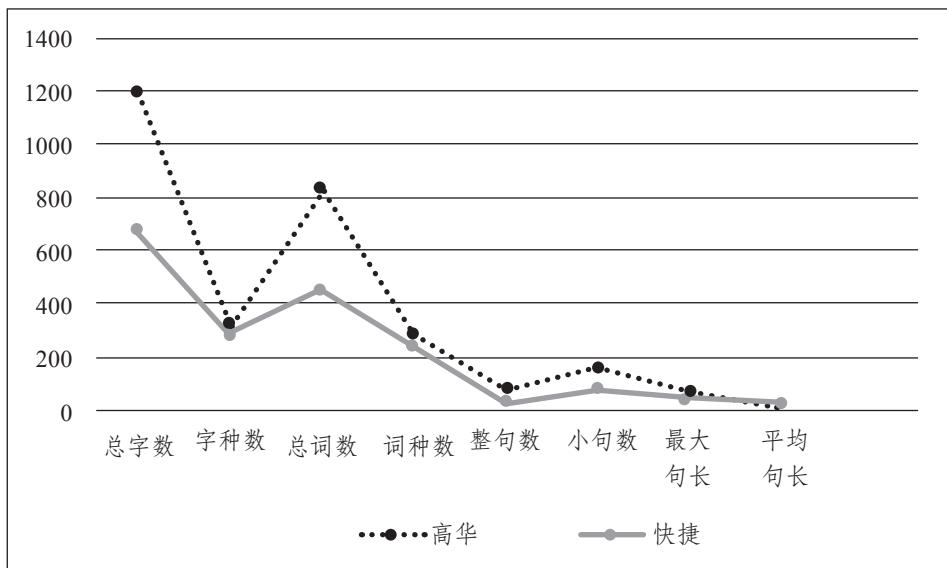
中学高级华文 (Higher Chinese) 课程是为华文程度较高的中学生设立的。据此, 高级华文的课文水平就应该比快捷华文来得高。但是, 应该高出多少, 课程标准并没有明确指定。如表 5 所示, 对比这两种课程的课文发现, 和字词有关的语文特质差异显然较大, 而和句子有关的则较小。

图 5 展示了高级华文和快捷华文四个年级课文的语文特质平均数。结果表明, 和字词有关的语文特质, 两种课程课文的总字数和总词数都有明显的差距, 但是, 字种数和词种数的差异比较小。最明显的差距出现在中二课文的总字数 (900)、总词数 (630)、和小句数 (112) 上, 效果强度分别为 5.2、5.2、6, 说明差异很大。如图 8 所示, 整体而言, 高级华文和快捷华文在语文特质方面的差异, 主要体现在总字数和总词数两个方面。

表 5: 高级华文与快捷华文的语文特质与可读性差异

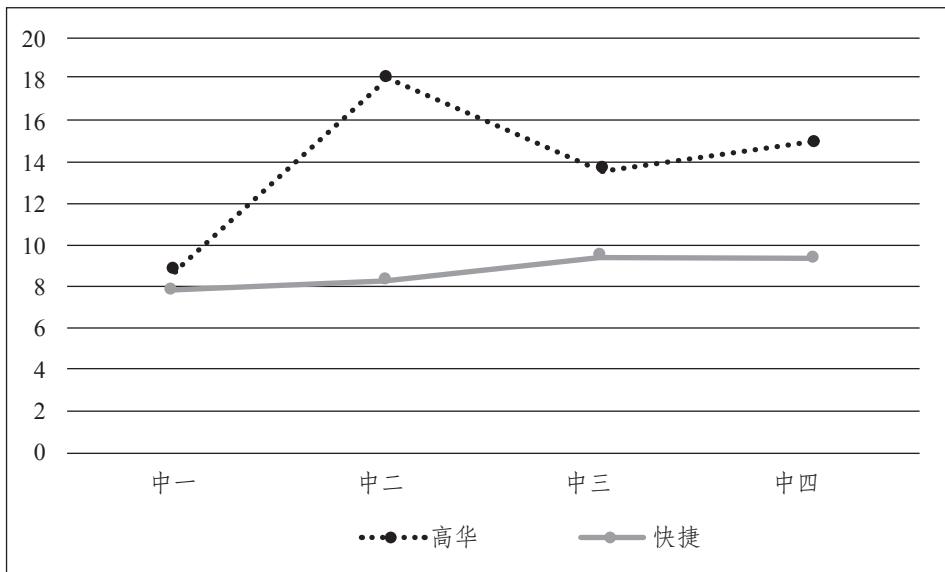
	总字数	字种数	总词数	词种数	整句数	小句数	最大句长	平均句长	可读性
高华中一	748.8	291.8	532.0	266.3	45.5	96.8	61.5	16.9	8.8
快捷中一	601.0	253.0	403.6	224.3	29.3	73.0	41.7	21.6	7.1
差数	-147.8	-38.8	-128.4	-42.0	-16.2	-23.8	-19.8	4.7	1.0
效果强度	2.0	1.5	1.9	1.6	1.8	1.3	3.8	-1.1	1.6
高华中二	1532.0	309.0	1057.0	268.3	81.5	185.0	56.8	19.6	18.0
快捷中二	632.0	281.3	427.1	246.1	28.6	73.1	52.3	23.0	8.3
差数	-900.0	-27.7	-629.9	-22.1	-52.9	-111.9	-4.5	3.4	0.7
效果强度	7.9	1.0	6.5	0.6	5.6	4.6	0.5	-0.9	12.1
高华中三	1188.5	324.5	821.3	302.0	61.8	165.3	56.8	19.2	13.6
快捷中三	719.4	304.3	478.3	266.7	28.7	86.1	52.6	25.3	8.5
差数	-469.1	-20.2	-343.0	-35.3	-33.0	-79.1	-4.2	6.1	-5.2
效果强度	4.6	0.6	4	0.9	5.8	5.2	0.6	-2.5	4.7
高华中四	1269.0	333.5	887.8	289.3	84.5	167.3	59.5	17.3	14.9
快捷中四	716.0	291.0	467.0	258.3	28.9	78.9	49.1	25.8	8.3
差数	-553.0	-42.5	-420.8	-31.0	-55.6	-88.4	-10.4	8.5	-6.6
效果强度	5.2	1	5.2	0.7	6.9	6	1.4	-1.7	4.7

图 8. 高级华文与快捷华文的语文特质平均差异



如果将高级华文当作华文第一语文，根据新加坡的一般假设，第一语文比第二语文高出两个年级，那么，高级华文的课文应该比快捷华文的只高出两个年级。然而研究结果表明，在课文的可读性方面（图 9），高级华文中一比快捷华文中一的课文高出一学年（8 至 9 年级），而高级华文中二却比快捷华文中二的课文高出十年（8 至 18 年）。这个差距到了中三和中四，缩回到 4 年 2 个月和 5 年 7 个月。由此得出高级华文课文的可读性与快捷华文课文的可读性之间的差异并不稳定，并且不太符合新加坡的一般假设。

图 9. 高级华文与快捷华文的可读性



结论与建议

综合以上分析结果，可归纳以下四点：

1. 小学华文课文各语文特质有适当地随年级而变化，逐年增加，惟小六的句子变化例外。可读性方面，进度大体正常，不过小二至小三和小四至小五的进度略高。
2. 中学快捷华文的语文特质变化不大，主要在总字数和总词数方面，小句数在中二到中三的变化比较大，但是从中三到中四反而减低。可读性方面，中一到中四只相差略多于一年。
3. 中学特别课文的语文特质和可读性，变化的情况不符初衷。中一到中二的变化正常，而中二则和快捷课程相同。特别需要注意的是，在中三和中四，特别课程的课文反而比快捷课程的难。
4. 中学高级华文课文和快捷课程课文的差异，主要在总字数和总词数方面。可读性方面，两种课程相应年级之间的差异并不一致，其中中二的差距最大。

整体来说，各课程课文的差异主要在和字词有关的语文特质，而

和句子有关的语文特质差异甚小。这意味课文的可读性，受字词的影响远高过受句子的影响。简而言之，课文的篇幅是可读性的关键因素。

基于以上分析，本研究提出以下建议：

1. 课文的深浅，除了篇幅的影响之外，句子的长短应该起比较大的作用。虽然字词有随年级增加的必要，但是，句子的量（句长）和质（句型）也应该随年级增长而变得比较复杂。

2. 各课程之间在语文特质和可读性两方面的差异应该配合课程具体的目标，在相应年级之间保持比较一致的距离，才能实现课程应有的功能。

3. 新编写的课文，在正式推出之前，应该加以检查，力求其语文特质和可读性符合课程的对象和教学的需要。

语文课文是语文教学成败的关键因素，而且课文的编写费时费力。如果课文不符合课程的目标和教学的需要，不但徒劳无功，而且会造成不必要的困扰，甚至带来误会。例如因为课文太难导致学生成绩不理想，却被误解为教不得法，学不得力。有科学化的可读性公式，能客观地、有效地检查新编或新选课文的语文特质和可读性，进而提高课文的适用性，有利于提高教学效率。

参考资料

- 林进展、赵春生、洪瑞春、吴福煥、王志豪 (2015) 基于语料库的新加坡华文教学资源平台开发。《华文学刊》，13 (1)， 1-15。
- 孙汗银 (1992) 《中文易懂性公式》。北京师范大学硕士学位论文。
- 宋曜廷、陈茹玲、李宜宪、查日龢、曾厚強、林維駿、張道行、張國恩 (2013) 中文文本可读性探討：指标选取、模型建立与效度验证。《中华心理学刊》，55 (1)， 75-106. DOI: 10.6129/CJP.20120621
- 王蕾 (2008) 可读性公式的内涵及研究范式—兼议对外汉语可读性公式的研究任务。《语言教学与研究》，第 6 期，46-53。<http://yyjx.cbpt.cnki.net/WKB3/WebPublication/paperDigest.aspx?paperID=2EAC9694-FD8D-4D02-B341-D9E7AEEFE15A>
- Bravos, K. (January 14, 2010). Readability Tests and formulas. Ideosity. <http://www.ideosity.com/ourblog/post/ideosphere-blog/2010/01/14/readability-tests-and-formulas>
- Common Core State Standards Initiative (2019). *English Language Arts Standards » Standard 10: Range, Quality, & Complexity » Measuring Text Complexity: Three Factors*. <http://www.corestandards.org/ELA-Literacy/standard-10-range-quality-complexity/measuring-text-complexity-three-factors/>
- Yang, S. J. (1971). *A Readability Formula for Chinese Language*. Ph.D. dissertation, University of Wisconsin.

Language Features and Readabilities of Singapore's Chinese Language Texts for Primary and Secondary Schools

SOH, Kay Cheng

Abstract

Language features and readability of Chinese texts are the essential conditions for the success of Chinese teaching. There are many English readability formulas, and they are often used to test the difficulty of English texts. Although there are a few readability formulas, they are appropriate for analyzing texts targeting at Chinese native speakers. This article first analyzes the language features of Chinese texts for primary schools in Singapore and then uses a newly developed Chinese as a second language readability formula to evaluate the difficulty levels of the texts, and further compares the language features and readability of Chinese textbooks for three streams of the secondary school. Results showed that text readability of primary textbooks was greatly affected by language features related to characters and words but not much by those related to sentences. Moreover, the readability of primary school texts increases by grade, but there is a lack of regularity in the texts for the secondary school.

Keywords: language features, readability formula, Chinese as a second language, regression analysis