Chapter 9
The Effect of Diglossia on Literacy in Arabic and Other Languages

John Myhill

Abstract This paper shows that basic literacy rates in Arabic-speaking countries are far lower than would be expected based upon their relative wealth, and argues that much of the explanation for this lies in their usage of a standard language which is based upon an earlier version of the language which no one speaks anymore—comparative evidence shows that languages of this type around the world consistently have uncommonly low literacy rates. The best policy for addressing this problem, so as to achieve a high rate of literacy while maintaining the traditional written language, would appear to be to use a strategy parallel to that adopted for languages such as Chinese, Japanese, and Sinhala: base early literacy, through the third or fourth grade, on written phonological representations of the different spoken dialects, and then switch to the traditional written language after this, when children are better able to deal with a writing system which is quite different from their own spoken languages.

Keywords Arabic · Diglossia · Language policy · Literacy · Mother tongue · Spoken language · Written language.

9.1 Introduction

One of the most obvious and striking features regarding Arabic is the remarkably low rate of basic literacy in Arabic-speaking countries. Consider the UNESCO data from 2007–2008 in the following table[1] (Table 9.1):

---

[1] Unless otherwise indicated, basic literacy data for individual countries which I will refer to in this study are from 2007–2008 and taken from http://en.wikipedia.org/wiki/List_of_countries_by_literacy_rate (based upon UNESCO data). Data on literacy and GDP per capital were not available

---

J. Myhill (✉)
Department of English Language and Literature,
University of Haifa, Mt. Carmel, 31905 Haifa, Israel
e-mail: john@research.haifa.ac.il

Table 9.1 Literacy ranking for Arabic-speaking countries (out of 180 countries)

Country	Literacy rate (%)	Literacy rank
Kuwait	94.5	76
Qatar	93.1	84
Jordan	91.1	92
UAE	90.0	98
Lebanon	89.6	101
Bahrain	88.8	104
Libya	86.8	113
Saudi Arabia	85.0	116
Syria	83.1	119
Oman	81.4	121
Tunisia	77.7	128
Algeria	75.4	132
Egypt	66.4	148
Sudan	60.9	156
Yemen	58.9	158
Morocco	55.6	162
Total	70.5	

Table 9.2 National literacy rates according to income of each state

Income level	Literacy rate (%)
High income	99.0
Middle income	89.9
Low income	60.2

We see here that of the 16 Arabic-speaking countries for which there are data, 14 have literacy rates which are lower than the average for the 180 countries listed, and even the two exceptions, Kuwait and Qatar, are barely above average.

Such data are alarming. But the literacy situation in Arabic-speaking countries is even worse if we consider how much money these countries have available to support education. Wealthier countries have more money to spend on education and can thus be generally expected to have higher literacy rates, and this is clearly supported by UNESCO data presented below in Table 9.2:[2]

Given the general association of income with literacy rates illustrated in Table 9.2, it might be thought that the low literacy rate in Arabic-speaking countries is the result of relative poverty in these countries. But this is not the case—in fact, the 16 Arabic-speaking countries taken together rank above average in terms of GDP per capital, with an average ranking of 72 out of 182 countries. Nonetheless, their

---

for Iraq and the Palestinian territories. The data from Sudan are only from northern Sudan; southern Sudan is entirely non-Arabic speaking and has been in a state of almost constant war against the north for the last 50 years, so presumably the literacy rates are lower there. The overall literacy rate of 70.5% for Arab countries is lower than the median for the countries because there is a strong tendency for the Arab countries with the highest literacy rates to have the smallest populations.

[2] Unless otherwise indicated, GDP per capita data which I will refer to in this study are from 2009 and are taken from http://en.wikipedia.org/wiki/List_of_countries_by_GDP_%28nominal%29_per_capita.

Table 9.3 Literacy ranking and GDP per capita ranking for Arabic-speaking countries (out of 180/182 countries)

Country	Literacy rate (%)	Literacy rank	GDP/capita	GPD/cap rank	GDP rank -lit. rank	Expected lit. rate (%)	Lit.rate- exp. lit. rate (%)
UAE	90.0	98	$ 45,615	9	−89	99.7	−9.7
Qatar	93.1	84	$ 59,990	4	−80	99.8	−5.7
Oman	81.4	121	$ 15,996	39	−82	99.0	−17.6
S. Arabia	85.0	116	$ 14,745	42	−74	99.0	−14.0
Bahrain	88.8	104	$ 19,817	33	−71	99.0	−10.2
Libya	86.8	113	$ 9,511	54	−59	97.8	−11.0
Morocco	55.6	162	$ 2,882	108	−54	88.0	−32.4
Kuwait	94.5	76	$ 27,835	26	−50	99.0	−4.5
Lebanon	89.6	101	$ 8,951	57	−44	97.6	−8.0
Algeria	75.4	132	$ 3,996	95	−37	90.7	−15.3
Tunisia	77.7	128	$ 4,171	92	−36	91.5	−13.8
Egypt	66.4	148	$ 2,450	116	−32	85.5	−19.1
Sudan	60.9	156	$ 1,397	128	−28	77.7	−16.8
Yemen	58.9	158	$ 1,061	140	−18	71.8	−12.9
Syria	83.1	119	$ 2,615	112	−7	87.0	−3.9
Jordan	91.1	92	$ 4,199	91	−1	91.2	−0.1
Average		*119*	–	*72*	*− 47*		*− 12.2*

literacy rates are **still** remarkably low. Table 9.3 shows the literacy ranking and GDP per capita ranking for Arabic-speaking countries.

Table 9.3 summarizes data from the 16 Arab states for which data are available for literacy rates (the first column), GDP per capita (the third column), ranking in these two categories out of the 180 and 182 countries listed (the second and fourth columns), GDP rank minus literacy rank (the fifth column, e.g. for the UAE 9−98=−89), the expected literacy rate based upon GPD per capita (the sixth column, for example, as the world's fourth wealthiest country Qatar would be expected to have a literacy rate of 99.8 %, equal to that of Latvia, the country with the fourth highest literacy rate), and the real literacy rate minus the expected literacy rate (the seventh column, for example, for Morocco 55.6 %−88.0 %=−32.4 %).

As Table 9.3 shows, every single Arab state has a lower literacy rate than would be expected given its per capita GDP—more than that, it is generally **much** lower, an average of 47 places out of 180 on the world scale and 12.2 %. The situation looks worse in different countries depending upon which measure is used. In general, the wealthier Arab countries are worse in terms of relative ranking (in fact the top five countries in Table 9.3 have the five lowest GDP-rank-minus-literacy ranks in the world), while the poorer Arab countries are worse in terms of comparison between actual and expected literacy rates (Oman and Morocco are extremely bad no matter which criterion is used), but with the exception of Syria and Jordan the situation is quite bad everywhere however this is calculated.

Table 9.4 Percentage of
GDP spent on primary
education

Region	GDP (%)
North America and Western Europe	21.8
Central and Eastern Europe	16.6
East Asia and the Pacific	14.6
Arab states	13.5
Sub-Saharan Africa	12.8
Latin American and the Caribbean	12.6
South and West Asia	9.7
Central Asia	9.3

The unfortunate situation described above cannot be attributed to low funding for primary education in Arab countries. In fact, the Arab states taken as a whole are slightly **above** average in terms of the percentage of GDP spent on primary education, as shown in Table 9.4 above:[3]

Nor is evidence for the literacy problem in Arabic-speaking countries limited to shockingly low literacy rates only. In an article published in the Journal of Higher Education entitled 'The Arabic publishing scene is a desert, critics say', Del Castillo (2001) writes:

> … the quantity of books published in the Arab world is small, especially relative to the region's population. There are 275 million Arabic speakers in 22 countries, but for Middle Eastern publishers, print runs of 5,000 are considered huge. (p. 55)

In the same vein, in her summary article on literacy in the Arab world for 'The Cambridge Handbook of Literacy', Haeri (2009, p. 423) observes that:

> All available statistics on literature and book-reading point to the generality of the conclusion that the public educational systems in most countries in the Arab world produce graduates or dropouts who do not like to read or write beyond a minimum that is required of them. (p. 423)

What is the cause of this literacy problem in the Arab world? We really do not know (see Saiegh-Haddad and Spolsky, in this collection). In fact, as noted by Haeri (2009), there have been almost no ethnographic studies of how Arabic literacy is taught (or fails to be taught) in Arabic-speaking countries, the only such study being Wagner (1993). As Wagner and Haeri observe, social factors such as attitudes towards literacy and uninspiring teaching methodology may certainly play a role (see Rosenhouse, in this collection). But anyone who has much experience with school systems around the world will know that Arab countries are not at all unusual in this respect—in fact such a situation is quite normal in many if not most countries, but rarely is the literacy situation as dismal as it is in Arab countries, so this cannot be the main cause of the problem.[4] It makes sense, then, to ask whether the explanation

---

[3] The data in Table 9.4 are from 2005 and taken from http://www.uis.unesco.org/template/pdf/EducGeneral/Factsheet07_No6_EN.pdf.

[4] The unusually low literacy rates in Arab countries do not appear to be the result of distinctive cultural biases discouraging female literacy. Arab countries show male/female literacy differentials which are typical of countries around the world, with substantially lower female literacy in

could lie in what Arabic-speaking students are being asked to learn, that is, the written form of the Arabic language itself. In this vein, on the basis of her own observations in Egypt, Haeri (2009, pp. 422–423) notes that studying the Arabic language generally turns Arabic speakers off to reading:

> … the majority of students feel alienated from such [Arabic language] classes and tell countless jokes about the language and teachers of grammar … [T]hey as well as older adults who were interviewed about their school experience stated that they found Arabic language classes extremely boring and unbearable—some even said that they hated these classes … A striking comment made both by high school students and older adults, men and women, was that they grew to dislike reading in general, especially 'longer pieces' like books. This was true even for the librarians that I interviewed. With few exceptions, people educated in public schools stated that they find the language of books too difficult and it takes them too long just to read a few pages. For fiction and nonfiction reading material, they commented that they found the language 'heavy' and 'scary' and that they simply did not enjoy the activity. [pp. 422–423; emphasis in original].

What it is it about the Arabic language that discourages reading and hinders literacy? A number of factors related to the Arabic script have been identified by researchers which may contribute to literacy problems in Arabic. These include in particular the facts that (1) many of the letters have a variety of different forms, and (2) diacritics are used extensively and in a manner which is phonologically inconsistent (see e.g. Azzam 1984; Bentin and Ibrahim 1996; and Ibrahim et al. 2002, 2007 for comparison with Hebrew; Frost et al. (1987) for comparison with English and Serbo-Croatian; and Roman and Pavard 1987 for comparison with French).

In the present paper, however, I would like to consider the possible effect of a different factor. This is the radical difference between on the one hand the established written language which is taught in school in Arab countries and upon which literacy acquisition and evaluation are based, and on the other hand the spoken dialects which children acquire before coming to school and speak in their day-to-day life (see Saiegh-Haddad and Henkin-Roitfarb, in this collection). This difference is one example of a situation which linguists have referred to as *diglossia* (see e.g. Ferguson 1959, 1991; Wexler 1971; Fellman 1975; Eckert 1980; Scotton 1986; Berger 1990; Daltas 1993; Schiffman 1997; Hudson 2002; Saiegh-Haddad 2003, 2004, 2005, 2007, 2012; Saiegh-Haddad et al. 2011; Khamis-Dakwar 2005, 2007). In a diglossic situation, the spoken language in a community, referred to by linguists as L (for 'low'), differs significantly from the community's written language, known as H (for 'high'); in some cases L may have some limited written usages (e.g. for folk poetry, songs, children's books, etc.), while conversely H may have some spoken usages (e.g. in television news, in speeches and religious sermons, or in the language in which teachers speak to students, etc.), but the general division of functions is clear.

Arabic was one of the four exemplary cases of diglossia discussed in Ferguson's foundational article and it has remained one of the cardinal exemplars of this

---

countries with generally low literacy rates but the gap being narrowed or even eliminated in countries with higher literacy rates—for example, for the four Arab countries with a literacy rate of at least 90 %, the average male–female difference is only 2 %.

phenomenon. In popular Arabic usage (which I will follow in the present paper), the H language is referred to as *fusħa* while the various dialects of L (which differ from each other enough so as to be in many cases mutually unintelligible) are referred to as *ʕa:mmiyya*. Fusħa and the various *ʕa:mmiyya*s differ significantly in a number of linguistic features, including phonological inventory of both consonants and vowels, phonotactics, agreement patterns, grammatical case, vocabulary, basic word order, etc.

Given the diglossic situation described above, Arabic-speaking children who are learning to read in primary school are thus confronted with a task which is quite different from the one encountered by their peers who speak and learn to read, for example, Hebrew or English. We will see that facts related to literacy in languages around the world give considerable reason to believe that such a radical difference between the child's spoken language and the language s/he learns to read and write in primary school results **in every other case** in serious problems in acquiring literacy, and so it stands to reason that this is true for Arabic as well (this does not mean, however, that the diglossic situation is the **only** source of the literacy problems being experienced in Arabic-speaking countries).

It should be made clear that this study is based upon **basic** literacy rates, associated with people 'who can with understanding both read and write a short simple statement on his everyday life', rather than **functional** literacy rates, associated with someone 'who can engage in all those activities in which literacy is required for effective functioning of his group and community and also for enabling him to continue to use reading, writing, and calculation for his own and the community's development' (from UNESCO's *Revised Recommendation concerning the International Standardization of Educational Statistics*; see e.g. Gray 1956; Levine 1994; Verhoeven 1994, 1997). Basic literacy rates are measured somewhat differently in different countries, so that simple comparison can in certain cases be misleading. My general approach to dealing with this limitation has been to simply gather and report data from as many countries as possible and to focus upon conclusions which are supported by a huge amount of clear data from a relatively large number of languages. For example, the conclusion that there is a serious problem with literacy in Arabic-speaking countries is based upon data from 16 different countries in **all** of which the literacy rates are lower than expected and in **most** of which they are **much** lower than expected. Most of my discussion here will be based upon data of this type (although there will be a few cases in which I will make speculations based upon less extensive data, and this will be made clear to the reader).

The 'basic literacy' which is measured by basic literacy tests is very basic indeed—just the ability to understand a 'short simple statement'. What is really more important to a person's ability to contribute productively to society is **functional** literacy, and it would of course be preferable to compare data from different countries on functional literacy rather than basic literacy, but unfortunately such data do not exist, and furthermore the criteria for calculating functional literacy differ even more radically from one country to another than do the criteria for calculating basic literacy, so that in doing a broad comparative study there is no choice but to use basic literacy data. It can be assumed that under normal circumstances the number of

functionally illiterate people in a given country will be far greater than the number of people who cannot pass a basic literacy test. For example, Doets (1994) estimates that even though the basic literacy rate in Holland is 99+%, functional illiteracy among native speakers may be as high as 18%, depending upon how this is measured. Given that the overall basic literacy rate in Arab countries is only 70.5%, this means that functional literacy is really a very serious problem in Arabic countries in terms of the implications it might have on the employability and productivity of the population.[5]

Section 9.1 of this paper will present evidence suggesting that education in the mother tongue—at least in the first few years of schooling—is the most effective way to teach literacy. In Sect. 9.1.2, we will see that a certain type of diglossic situation, of which Arabic is one example, in which the H language is based on usage from hundreds of years ago and not remotely corresponding to anyone's everyday spoken language today, is particularly problematic in terms of literacy. This gives evidence that at least part of the literacy deficit in Arab countries may be due to this type of diglossia, which is supported by data from studies such as Eviatar and Ibrahim (in this volume, see also references therein), which have shown that from a cognitive perspective *Fusħa* is effectively a non-native language for Arabic speakers today, even though it is popularly considered to be 'the same language' as *ʕa:mmiyya*. Section 9.1.3 then discusses how literacy problems in Arabic-speaking countries resulting from this situation might productively be addressed.

9.1.1   The Importance of Education in the Mother Tongue

One of the most important issues affecting language policy is the connection between the acquisition of literacy and the relationship between the spoken language of the child and the written language which s/he is learning. It has been recognized for some time that it is problematic for a child to begin to learn to read and write using a written form which is understood to be a different language from the language which s/he has grown up speaking (see Saiegh-Haddad and Spolsky, in this collection). Thus already in the 1950s UNESCO observed that:

---

[5] Available comparative literacy data refer specifically to the **attainment** of literacy by **adults**, which is not the same thing as the **acquisition** of literacy by **children**. It would obviously be preferable for the purposes of the present study to rely upon the latter type of data, but unfortunately, comparable data of this type from a wide variety of languages do not exist. In such a situation the best that can be done is to assume that the correlations which are found between language policy and adult literacy data reflect the effect which these policies have upon the acquisition of literacy by children, particularly if a plausible account can be given to explain these correlations. The distinction between data on adults' attainment and data on children's acquisition is particularly problematic in countries in which there are a significant number of immigrants who are not native speakers of the national language. In practice, however, this phenomenon is almost entirely restricted to Western European and Anglophone states in which the basic literacy rate is in any case assumed to be at least 99%.

> On educational grounds, we recommend that the use of the mother tongue be extended to as late a stage in education as possible. In particular, pupils should begin their schooling through the medium of the mother tongue because they understand it best and because to begin their school life in the mother tongue will make the break between the home and the school as small as possible. (UNESCO 1953, pp. 47–48)

This position has been supported by numerous empirical studies (see e.g. Gudschinsky 1977; Okedara and Okedara 1992; Dutcher and Tucker 1997; Mehrotra 1998, etc.).

I will add to these studies further evidence taken from the language policies of the countries which were colonized by Great Britain and France and then became independent after the Second World War (see discussion of language policies in these countries, both before and after colonization, in e.g. Calvet 1974; Bokamba 1984; Phillipson 1992; Dumont and Maurer 1995; Alidou 1996; Fishman et al. 1996; Bokamba and Tlou 1997; Gill 1999; Powell 2002; and Salhi 2002). We can divide these countries into three groups: (1) those which no longer use the colonial language but have rather developed an indigenous language as their official language, (2) those which use both the colonial language and one or more indigenous languages as official, and (3) those which only use colonial languages as official. It turns out there is a very strong pattern of literacy rates being highest in ex-colonies which have entirely rejected the colonial language as official, while literacy rates are lowest in those ex-colonies which continue to use only the ex-colonial language as official while not giving this status to indigenous languages. This is shown in Tables 9.5, 9.6, and 9.7 on the following pages.[6]

When indigenous languages have official status, they will naturally be used in schooling from the earliest age, and this will mean that children will generally be educated in, and be taught how to read from the beginning in, their native language (although this will not invariably be the case, as in some cases they may be speakers of a non-official indigenous language which is not used in education), and this is presumably why the literacy rates are highest in countries which have rejected the colonial language. On the other hand, when only foreign ex-colonial languages are official, children will generally not be educated in their mother tongue (although there may be cases in which non-official languages are used for some limited educational purposes). We see in Tables 9.5, 9.6, and 9.7 that there is an extremely strong tendency for ex-colonial countries which use an indigenous language as their official language to have much higher literacy rates than ex-colonial countries which use English and/or French as their official language.

Furthermore, a number of the exceptions to this general pattern have obvious explanations. For example, in Bangladesh, where the literacy rate is only 53.5 % even though the only official language, Bengali, is indigenous, written Bengali is nevertheless radically different from spoken Bengali, a diglossic situation parallel to that of Arabic, which means that the language which young children learn to read and write is still very different from the language they have already learned how

---

[6] I have excluded from Tables 9.5, 9.6 and 9.7 countries in which the majority of people speak English, French, and Arabic and/or which use Arabic as their official language.

Table 9.5 Literacy rates in ex-colonies of Great Britain and France which only have indigenous languages as official

Country	Literacy rate (%)
Cyprus	97.7
Israel	97.1
Maldives	97.0
Brunei	94.9
Myanmar	91.9
Sri Lanka	90.8
Vietnam	90.3
Cambodia	76.3
Laos	68.7
Nepal	56.5
Bangladesh	53.5
Average	*83.2*

Table 9.6 Literacy rates in ex-colonies of Great Britain and France which have English and/or French but also indigenous languages as official

Country	Literacy rate (%)
Singapore (E)	94.4
Fiji (E)	94.4
Malta (E)	92.4
South Africa (E)	88.0
Botswana (E)	82.9
Lesotho (E)	82.2
Swaziland (E)	79.6
Vanuatu (EF)	78.1
Comoros (F)	75.1
Uganda (E)	73.6
Kenya (E)	73.6
Tanzania (E)	72.3
Malawi (E)	71.8
Madagascar (F)	70.7
India (E)	66.0
Papua NG (E)	57.8
Pakistan (E)	54.2
CAR (F)	48.6
Average	*75.4*

to speak, so that the same explanation for low literacy rates which we will see in Arabic-speaking countries also accounts for the low literacy rate of Bangladesh. A similar account may be given for the low literacy rate in Pakistan: the indigenous language used in education, Urdu, is only spoken natively by 8 % of the population of the country, so that for the overwhelming majority of Pakistanis, primary education requires children trying to become literate in a foreign language, a task which barely half of them succeed in accomplishing.

A radically different colonial language policy was practiced by the leaders of the Soviet Union who inherited the territories which the Russian Empire had accumulated in the preceding centuries. These territories were occupied by speakers of a great variety of languages, the overwhelming majority of which had never or almost

Table 9.7 Literacy rates in ex-colonies of Great Britain and France which only have English and/or French as official

Country	Literacy rate (%)
Zimbabwe (E)	91.2
Gabon (F)	86.2
Rep. Congo (F)	81.1
Solomon Isl. (E)	76.6
Belize (E)	75.1
Nigeria (E)	72.0
Zambia (E)	70.6
Cameroon (EF)	67.9
Ghana (E)	65.0
Togo (F)	53.2
Ivory Coast (F)	48.7
Gambia (E)	42.5
Senegal (F)	41.9
Benin (F)	40.4
Sierra Leone (E)	38.1
Guinea (F)	29.5
Burkina Faso (F)	28.7
Niger (F)	28.7
Mali (F)	26.2
Average	*56.0*

never been written before. Bolshevik language policy was focused upon making everyone in the country literate as quickly and efficiently as possible, because of both general egalitarian ideals and a specific desire to have everyone read ideological writings. The leaders recognized that the most effective way to do this was by teaching everyone to read and write a written version of the language or dialect which they already spoke—even though in almost all cases this meant sending linguists into the field to develop new writing systems for languages which had more or less never been written before (Ornstein 1968; Lewis 1972; Pool 1978; Azrael 1978; Simon 1991).

In cases in which there were two or more related but distinct dialects, different writing systems were developed for each dialect, thereby turning them into separate languages and minimizing the difference between the spoken and written language in every case. Thus, the Turkic languages Kazakh, Kyrgyz, Uzbek, Turkmen, and Azeri were distinguished from each other, as were the Slavic languages Russian, Ukrainian, and Belarusian and the Baltic languages Lithuanian and Latvian. This is the reverse of the situation of Arabic, where a single written language is used to represent a huge variety of very distinct spoken dialects, in many cases mutually unintelligible. Soviet language policy also called for written languages based upon local dialects to be used in cases in which there was a related but distinct language with an already-established writing system in use outside of the Soviet Union. Thus written Moldovan was used instead of the related Romanian, written Tajik was used instead of the related Persian, written Estonian was used instead of the related Finnish, and the written Turkic languages mentioned above were used instead of the related Turkish.

Table 9.8 Literacy and GDP per capita for ex-Soviet Republics (out of 180 countries for literacy and 182 countries for GDP per capita)

Country	Literacy rate (%)	Literacy rank	GDP/capita	GDP/capita rank	GDP-lit
Tajikistan	99.6	10	$ 667	153	143
Kyrgyzstan	99.3	17	$ 851	147	130
Georgia	100.0	1	$ 2,450	117	116
Moldova	99.2	19	$ 1,514	127	108
Armenia	99.7	7	$ 2,615	113	106
Ukraine	99.7	7	$ 2,569	114	107
Turkmenistan	99.5	13	$ 3,451	101	88
Belarus	99.7	7	$ 5,166	80	73
Azerbaijan	99.5	13	$ 4,798	83	70
Uzbekistan	96.9	63	$ 1,176	133	70
Kazakhstan	99.6	10	$ 6,930	67	57
Russia	99.5	13	$ 8,681	59	46
Latvia	99.8	3	$ 11,466	47	44
Lithuania	99.7	7	$ 11,115	50	43
Estonia	99.8	3	$ 14,402	43	40
Average		*13*		*96*	*83*

At least for the 15 Union Republics, the language of the republic continued to be the language of primary education and also usually secondary education throughout the Soviet period. When the republics became independent in 1991, these languages became their respective languages of education. The results of this policy are shown Table 9.8, which gives data for the ex-Soviet Republics in 2007:[7]

As can be seen, the policy of educating people in written languages based directly upon their spoken usage has been astonishingly successful, producing basically universal literacy, even in countries which have very little money to spend on education, with literacy rankings which are on the average 83 points higher than what would be expected considering their GDPs per capita (and as we will see later in this paper, the only ex-Soviet state with a literacy rate lower than 99.5 %,

---

[7] It should be noted that in the overwhelming majority of cases, the numbers in Table 9.8 indicate literacy in the language of the state rather than in Russian. There is no reason to suspect that Russians are any more literate than are non-Russians in ex-Soviet states; indeed, the three republics with the highest proportion of ethnic Russians—Kazakhstan, Latvia, and Estonia—all have **higher** literacy rates than does Russia itself. It is possible that the government of the Soviet Union invested a relatively high proportion of their resources in basic education and that this would result in a relatively high rate of literacy compared to GDP per capita. While this hypothesis is certainly worth investigating, it should be pointed out that the data in Table 9.8 are from 2007, 16 years after the dissolution of the communist government of the Soviet Union, that I do not know of evidence that the Soviet Union spent a high proportion of its resources on education, and that in fact at present the countries listed in Table 9.8 are if anything spending a disproportionately **low** percentage of their GDP per capita on education (see http://www.nationmaster.com/graph/edu_edu_spe-education-spending-of-gdp, which has data for all of these countries other than Uzbekistan and Turkmenistan), averaging only 4.2 % and a ranking of 76 out of 132 countries, making their high literacy rates even more impressive.

Table 9.9 Literacy percentages for Soviet republics whose languages had no previous literary tradition

Country	1897	1926	1939	1959
Azerbaijan S.S.R.	9.2%	28.2%	82.8%	97.3%
Kazakh S.S.R.	8.1%	25.2%	83.6%	96.9%
Kyrgyz S.S.R.	3.1%	16.5%	79.8%	98.0%
Moldavan S.S.R.	22.2%	(n.d.)	45.9%	97.8%
Tajik S.S.R.	2.3%	3.8%	82.8%	96.2%
Uzbek S.S.R.	3.6%	11.6%	78.7%	98.1%
Turkmen S.S.R.	7.8%	14.0%	77.7%	95.4%

Uzbekistan, is also the only one which did not really follow standard Soviet procedure in terms of defining a nationality and establishing a standard language). These patterns are not the result of these countries focusing whatsoever limited financial resources they have on education—on the contrary, as we have seen in Table 9.4, Central Asia, where most of the countries with the highest GDP-literacy ranking are located, has the lowest rate of per capita spending on primary education of any region in the world.

Additionally, many of these cases involved developing literacy from scratch in the last century, using written languages which were invented by linguists sent by the Soviet government to create them on the basis of local usage. Table 9.9 shows the remarkable success of Soviet language policy in increasing literacy in republics whose languages had no previous literary tradition:[8]

As can be seen, literacy rates had only increased moderately by 1926, as the policy emphasizing mother-tongue literacy had only begun to be implemented and had not yet substantially affected the older adult population (based on whom literacy rates are calculated) who had grown up before the Revolution, but already by 1939 dramatic increases had taken place and this trend was even stronger in 1959 (progress in Moldova was slower because it did not become a Union Republic until after the Second World War). It is quite striking to compare the great success of these literacy programs with the catastrophic failure of language policies in ex-British/French colonies demonstrated in Tables 9.6 and 9.7, which rely upon English and French.

There is strong evidence, then, that primary education in a native language is much more efficient in terms of literacy than is education in a non-native language. But what about diglossic situations, in which education is in a written language which is not socially or politically foreign but which is nevertheless quite different from the spoken language? In the following section, we will consider a wide variety of cases of this type from around the world.

---

[8] The data in Table 9.12 are from Lewis (1972, p. 175). Data from 1897 are from the areas of the republics, which did not yet exist as political entities, and they are for languages other than those of the present-day republics, which had not yet been written.

9.1.2   Diglossia and Literacy

In the sense in which it was originally used in Ferguson (1959), the term *diglossia* refers specifically to the case in which H and L are understood to be forms of the same language which differ significantly one from the other and are used for complementary sets of social functions. Given this very general understanding, it is not clear **how** different the forms should be in order for the situation to be considered diglossia—presumably it would be reasonable to characterize different situations according to **degrees** of diglossia so that some are more diglossic while others are less diglossic, according to the magnitude of the difference between the H and the L.

Setting aside the question of degree of diglossia linguistic distance, there is reason to believe that there is a marked difference in the effects upon literacy rates of different **types** of diglossia, which I will enumerate in this section.

For one type of diglossia, the H language is the spoken language of people in a **different** country. Literacy data for this type of diglossia are given in Table 9.10.

The first six of these countries are in a very similar sociolinguistic situation: they are islands in the Caribbean where the people speak an English creole as their native language but the language of literacy is the same standard English which is used in schools in other countries (the United States or Great Britain), which is based on the native dialect of high-status people in these countries; I will refer to this type of diglossia as *external* diglossia.[9] The creoles are quite different from the standard language, which is essentially not spoken by anyone in these countries as a first language, so that these represent cases of diglossia in the sense that Ferguson originally used the term, but this in itself does not seem to be a great barrier to literacy—for the six countries, the average literacy rank is 57 while the average GDP per capita rank is 61. The situation in Switzerland is of the same general type—the languages of literacy (standard German, French, and Italian) are all based upon dialects spoken in other countries (Upper Saxon in Germany, Parisian in France, and Tuscan in Italian), and at least in the cases of German (constituting 75 % of the population of Switzerland) and Italian (constituting another 5 %), these are very different from the local versions. Switzerland is therefore also generally characterized by external diglossia—in fact, Swiss German was one of Ferguson's initial exemplary cases of diglossia—and Switzerland's GDP-lit of −30 might appear to suggest some negative effect of diglossia on literacy, but this is misleading: in fact, basic literacy in Switzerland is essentially universal, the rate of 99.0 % being conventionally assigned to Western countries with more or less universal literacy—given this convention, Switzerland's literacy ranking is in fact as high as it could be, even though this is not as high as its GDP per capita ranking. The cases listed in Table 9.10 suggest, then, that external diglossia is not in itself a great obstacle to literacy.

---

[9] The sociolinguistic situation on these islands is parallel to that of Haiti at the time of Ferguson's original article, when he used Haiti as one of his four exemplary cases of diglossia. Haiti differs in that the creole is French-based and the H was French. I have not included Haiti in Table 9.10 because the creole there has recently begun to be used as a language of education, so the situation is no longer diglossic in this sense.

Table 9.10 Literacy rates of countries in which the H language is based upon the everyday usage of spoken by people living in **another** country

Country	Literacy rate (%)	Literacy rank	GDP/capita	GDP/ capita rank	GDP-lit
Antigua	99.0	33	$ 13,150	45	+12
Barbados	99.7	6	$ 14,105	44	+38
Grenada	96.0	67	$ 5,969	73	−6
Jamaica	86.0	115	$ 4,684	84	−31
Saint Lucia	94.8	73	$ 5,671	76	+3
Trinidad	98.7	49	$ 15,000	41	−8
Switzerland	99.0	33	$ 63,536	3	−30

But there is a different type of diglossia, in which the H is not at present spoken as an everyday colloquial language by **anyone anywhere** but is rather based upon texts written in the fairly distant past which are understood to represent the 'correct' version of the language, on the basis of which present-day linguists have devised a standard language. We can call this *frozen* diglossia, in the sense that the written standard presumably represents a preserved version of the language as it was spoken long ago (the written language does continue to evolve, although in a manner more or less independent of the spoken language, and so I will use the term 'frozen' in scare quotes to emphasize that this is a technical usage). Arabic is one such language of this type, and we have already seen that literacy rates in Arab states are very low, particularly considering the relative wealth of these states. Table 9.11 on the following page gives literacy rates from a number of other cases of this type.

As can be seen in Table 9.11, the literacy rates for these cases of 'frozen' diglossia are generally quite low. Furthermore, they are even lower than would be expected given the relative wealth of these states. For instance, Iran is ranked 87th in the world in terms of GDP per capita but only 121st in terms of literacy rate, Bangladesh is 158th in terms of GDP per capita but only 164th in terms of literacy rate, and within India, Tamil Nadu, Andhra Pradesh, West Bengal, and Karnataka rank 8th, 11th, 10th, and 9th, respectively, in terms of GDP but only 11th, 13th, 16th, and 18th in terms of literacy rate.[10, 11] Thus, in almost every case of the 'frozen' type

---

[10] Per capita income data from Indian states are taken from http://en.wikipedia.org/wiki/List_of_Indian_states_by_GDP.

[11] Astonishingly, Wagner, Spratt, and Ezzaki come to a completely different conclusion, that "the findings support the proposition that children in certain social and linguistic contexts need not be taught in their mother tongue in order to achieve literacy norms of the majority language group" (p. 31)—that is, that children can be taught literacy just as efficiently in a second language as in their mother tongue. The authors seem to be under the impression that it would be satisfactory for Berber-speaking Moroccans 'to achieve literacy norms of the majority language group', that is, Arabic-speaking Moroccans, apparently having neglected to check what these norms actually are. In fact, the results of Morocco's literacy program for **Arabic** speakers are not merely unsatisfactory but catastrophic: the literacy rate in Morocco is only 55.6 %, **32.4 %** lower than what would be expected give the GDP per capita in the country—this is the third worst differential in the world, being exceeded only by Chad (64.6 %−31.8 %=32.8 %) and Mali (59.3 %−26.2 %=33.1 %), two countries which have only foreign languages as official—and it was undoubtedly even lower in 1989 when the article was written. This is presumably not due to the fact that 45 % of the

Table 9.11 Literacy rates for countries/states with 'frozen' diglossia

Country	Literacy rate (%)
Sri Lanka (Sinhala)	90.7
Iran (Persian)	82.3
Tamil Nadu (Tamil)	74.2
Andhra Pradesh (Telugu)	72.5
West Bengal (Bengali)	71.6
Karnataka (Kannada)	69.3
Bangladesh (Bengali)	53.5

The languages are given in parentheses; Tamil Nadu, Andhra Pradesh, West Bengal, and Karnataka are states in India. (Literacy data from Indian states are from http://en.wikipedia.org/wiki/Indian_states_ranking_by_literacy_rate)

of diglossia which Arabic also represents, the literacy rate is lower than would be expected considering the relative wealth of the state, although the differences in the cases listed in Table 9.11 are not nearly as great as in the Arab states. The one exception to this pattern is Sri Lanka, which ranks 121st in the world in per capita income but 94th in the world in literacy rate. This pattern is exactly the reverse of the pattern observed in all of the other states with 'frozen' diglossia. I will return to this one exceptional case later in this paper.

Further data in support of the conclusion that 'frozen' diglossia is a serious obstacle to literacy come from comparing Persian and Tajik, which are fairly closely related languages, belonging to the same branch of the Indo-Iranian family. Until the Soviet period, Tajik was understood to be a dialect of Persian and, to the extent that speakers of what is known today as Tajik were literate at all, they used Persian as their literary language. As we have seen, the Soviet government developed Tajik as a distinctive written language, based upon the spoken language of Tajikistan. Comparative literacy and GDP data are given for Tajikistan and Iran in Table 9.12:

As can be seen, Tajikistan's literacy rate is more than 17 % higher than that of Iran, even though the GDP per capita of Tajikistan is less than 1/6 that of Iran. Literacy in Tajikistan has obviously benefited enormously from developing and using a standard written language based directly on the local spoken language rather than the 'frozen' Persian standard.

Greek provides further evidence supporting the same conclusion. Until 1976, Greek was a case of 'frozen' diglossia, but since then the H language, Katharevousa, which was based upon the Byzantine language rather than on any group's contemporary spoken usage, has been replaced as the standard language by a written language based upon the spoken language, which is called Demotiki (meaning 'the

---

population of Morocco are native speakers of Berber rather than Arabic, because Wagner, Spratt, and Ezzaki report that there is no difference in reading ability between Berber speakers and Arabic speakers. The real situation is therefore not that Berber speakers do as **well** as Arabic speakers but rather that Arabic speakers do as **badly** as Berber speakers—being a native speaker of Arabic is not an advantage in learning to read Arabic. But this is only the situation because the traditional written language in Arabic is completely different from the spoken dialects.

Table 9.12 Literacy rates and GDP per capita in Tajikistan and Iran

Country	Literacy rate (%)	Literacy rank	GDP/capita	GDP/capita rank
Tajikistan	99.6	10	$ 667	153
Iran	82.3	121	$ 4,399	87

Table 9.13 Literacy rates in Greece with and without 'frozen' diglossia

Year	Literacy rate (%)
1971 (with 'frozen' diglossia)	86.0
2007 (without diglossia)	97.1

language of the people'; see Browning 1982; Frangoudaki 1992). After this was done, literacy rates increased drastically, as is shown in Table 9.13:

The pattern, then, is quite clear: 'frozen' diglossia appears to be associated with very adverse consequences for literacy, not just for Arabic but in general. However, this does not appear to be necessarily true of **external** diglossia, as we have seen in Table 9.10. Why is there this difference? There are a number of possible explanations. It may be the case that the difference between the everyday spoken language and the standard language is simply much greater in the cases of 'frozen' diglossia than in the cases of external diglossia (although it is not clear how this would be tested, as there are no established criteria for measuring linguistic distance) . Or it may be that when H is spoken in other countries as is the case in external diglossia, speakers of L can at least hear it used naturally in everyday speech when they interact with or overhear native speakers of H when one or the other is traveling, or on television or radio, so that their acquisition of H is facilitated by naturalistic exposure, whereas when H is never spoken as in 'frozen' diglossia this does not happen. Or it may be that 'frozen' diglossia Hs are characterized by many artificial grammatical rules which prescriptive linguists have devised but which are not natural for average people to learn or use, whereas the grammatical rules of the spoken languages in external diglossia Hs are for the most part based upon naturalistic everyday usage, albeit of people living in a different country, and this makes these languages easier to learn. Or it may be a combination of these factors.

To sum up, the comparative evidence clearly suggests that Arabic diglossia, as an instance of 'frozen' diglossia, poses a significant obstacle to literacy. This has in fact been suggested on the basis of language-internal evidence by studies by Abu-Rabia (2000), Maamouri (1998), and Saiegh-Haddad (2003, 2004, 2005, 2007, 2011). In fact, psycholinguistic research has demonstrated that from a cognitive perspective (though not a social or political one), Fusħa is a non-native language for native speakers of colloquial Arabic (see Eviatar and Ibrahim in this volume and the extensive research referred to there), so that in this sense the situation in Arabic-speaking countries is parallel to that of ex-colonial countries which use English or French as their only official language—it is then to be expected that the literacy situation in Arab countries would be more or less equally dismal. Haeri (2009, p. 420) states in her summary article on literacy in the Arab world that:

My central argument in this chapter is that the main reason for exceedingly low literacy in most of the Arab world is that the language of education in the public educational systems is Classical Arabic and modernized versions of it.

9.1.3   What is to be Done?

What then can be done about this situation, to improve the extremely low literacy rates of Arabic speakers? One possibility is to replace the current H language, based upon the spoken languages of pre-Islamic tribes as well as the Koran and the old Arabic literature, with a new H language based upon one of the spoken dialects. Presumably Egyptian Arabic would be the most reasonable choice, given the fact that (1) it is by far the most widely known of the present-day spoken dialects, particularly because of the popularity of Egyptian movies, and (2) it is relatively 'central' as a dialect, not having as many distinctive characteristics as dialects spoken farther to the west or farther to the east (other than the pronunciation of jeem as [g]). This approach might replace the present H of 'frozen' diglossia with an H that is spoken in Egypt and might turn the situation into external diglossia elsewhere. This might have a positive effect on literacy rates, as we have seen external Hs are much easier to learn than 'frozen' Hs.

There are, however, a number of problems with this approach. First, it would be very difficult to eliminate the currently-existing H, because of its enormous cultural and religious value. Second, it would in a significant sense privilege Egypt over other Arab states, which would run counter to the general egalitarian spirit of Arab nationalism. And third, although such an approach might result in a significantly higher literacy rate among Arabic speakers than the present situation, it would still be less than ideal in the sense that the great majority of Arabic speakers would still be educated in a standard form which is quite different from their own spoken dialect, and there is reason to believe that this would produce less than optimal results regarding literacy.

Why is this? There is evidence suggesting that basic literacy is most effectively learned not only in the children's native language but specifically in a written language which is as close as possible to the child's native **dialect**, and that it is more efficient in terms of literacy to divide up neighboring dialects into different languages and to devise different written languages for each one. Consider for example data from the Slavic languages. The Slavic-speaking peoples are divided up into 11 different language-based nationalities with each nationality living in an area within which a fairly narrow range of dialects are spoken. This means that written languages can be used which are very close to the spoken daily language of essentially the entire population. Literacy and GDP per capita data for the Slavic-speaking countries are presented in Table 9.14 on the following page.

We see here that not only for the three ex-Soviet republics but also for Slavic languages in general, literacy rates are a good deal higher than would be expected on the basis of the GDP per capita. This supports the idea that literacy is most

Table 9.14 Literacy and GDP per capita rankings for Slavic-speaking countries

Country	Literacy rate (%)	Literacy rank	GDP/capita	GDP/capita rank	GDP-lit
Ukraine	99.7	7	$ 2,569	114	107
Belarus	99.7	7	$ 5,166	80	73
Russia	99.5	13	$ 8,681	59	46
Poland	99.3	17	$ 11,302	49	32
Bosnia	96.7	64	$ 4,365	88	24
Macedonia	97.0	61	$ 4,546	85	24
Slovenia	99.7	7	$ 24,111	30	23
Bulgaria	98.3	51	$ 6,623	72	21
Serbia	96.4	66	$ 5,821	75	9
Slovakia	99.0	33	$ 16,282	38	5
Czech Republic	99.0	33	$ 18,256	35	2
Croatia	98.7	49	$ 15,284	40	−9
Average		*34*		*64*	*30*

efficiently taught in a written language which is as close as possible to the particular spoken dialect of the language learner.

Data on related languages from the ex-Soviet republics (see Table 9.8) support the same conclusion—literacy rates are extremely high for not only the Slavic languages Russian, Ukrainian, and Belarussian but also the Baltic languages Latvian and Lithuanian and the Turkic languages Kazakh, Kyrgyz, Turkmen, and Azeri. And, strikingly, they are significantly **lower** for Uzbek, which was the only language of a Union Republic for which the principle dividing distinct spoken forms into different languages was not followed. For various cultural and political reasons, having to do with pre-modern political borders and the idea that Uzbeks were understood to be inherently sedentary while Turkmen, Kazakhs, and Kyrgyz were understood to be inherently nomadic, the Soviet authorities drew the borders of Uzbekistan and defined the Uzbek language to include not only the highly distinctive Southeastern dialects which are completely different from Kazakh, Kyrgyz, and Turkmen, but also some Northwestern Turkic dialects which are linguistically much closer to Kazakh and Kyrgyz. This meant that Uzbek was dialectally split in a way that none of the other languages of Soviet republics were, and because Standard Uzbek is based upon the Southeastern dialect, this means that the many Uzbeks whose native dialect is Northwestern have to learn a written language in school which is very different from their own spoken language. This is presumably related to the fact that of the ex-Soviet republics only Uzbekistan has less than essentially universal literacy.

So would it then be reasonable to adopt such an approach to the Arabic dialects, developing different written languages for each? It seems that it would not, for both religious and national reasons, because it would mean both eliminating the classical language **and** linguistically dividing among people who are now considered to be speakers of different dialects of the same language and belonging to the same nationality.

Table 9.15 Literacy rates and GDP per capita for Sri Lanka and Arabic-speaking states with comparable GDP per capita

Country	Literacy rate (%)	Literacy ranking	GDP per capita	GDP ranking
Sri Lanka	90.8	94	$ 2,085	121
Morocco	55.6	162	$ 2,882	108
Syria	83.1	119	$ 2,615	112
Egypt	66.4	148	$ 2,450	116
Sudan	60.9	156	$ 1,397	128
Yemen	58.9	158	$ 1,061	140

There is, however, a way to get around this problem. This would be to use a written version of the local dialects specifically in **primary** education, for the first three or four grades, before switching to teaching the established literary language beginning in the fourth or fifth grade. The most striking evidence supporting this approach comes from the only country with a 'frozen' diglossia and hence a non-spoken H which has a literacy rate which is **higher** than what would be expected from its GDP per capita. This country is Sri Lanka, whose national language is Sinhala. It will be remembered that in every other case of a 'frozen' diglossia H—Arabic (in every Arabic-speaking country), Persian, Bengali (in both Bangladesh and the Indian state of West Bengal), Tamil, Telugu, and Kannada—the literacy rate is lower, and often **much** lower, than would be expected from the GDP per capita. In contrast, in Sri Lanka we see exactly the reverse pattern, with a literacy ranking of 94th but a GDP per capita ranking of 121st. Consider for example the data in Table 9.15, which compares literacy rates in Sri Lanka with those of Arabic-speaking countries with a GDP close to that of Sri Lanka (within 20 places on the GDP per capita ranking).

As we see in Table 9.15, Sri Lanka's literacy rate is 25 % higher than the average of these five Arab countries (65 %). This difference becomes even more striking when it is considered that about 20 % of the population of Sri Lanka speak Tamil as their native language, and although separate literacy figures are not available for the Tamils of Sri Lanka, it would be safe to assume that the Tamil speakers pull down the overall literacy rate, because the literacy rate in Tamil Nadu in India is only 74.2 %.

Why do we find in Sri Lanka a completely different literacy situation from what we observe in any other state with 'frozen' diglossia? This is a question that requires further investigation. However, there is one obvious thing distinguishing Sri Lanka from the other cases considered here and this is that **reading in Sinhala is taught in L for the first four years of school, with students only beginning to learn to read in H in the fifth grade**. This is not the case for **any** of the other languages with 'frozen' diglossia Hs, either the Arabic-speaking countries or any of the states listed in Table 9.11.

Striking as this finding may seem, it is exactly what would be expected if we assume both that basic literacy is best taught through the mother tongue—as is the assumption of UNESCO—and that the Hs in 'frozen' diglossia contexts are

cognitively non-native languages from the perspective of speakers of the associated Ls (see Eviatar and Ibrahim in this volume). Sinhala-speaking children acquire literacy effectively because they are initially educated in a written version of their own spoken language; thus the Sinhalese have managed to achieve relatively high literacy ranking while at the same time preserving their culturally valued 'frozen' H. In contrast, speakers of other languages with 'frozen' diglossia do not acquire literacy effectively because they are initially educated in the non-spoken 'frozen' H, which is cognitively a foreign language to them. The same sort of thinking lies behind teaching methodologies such as the Pitman Initial Teaching Alphabet (Downing and Latham 1967), which uses a modified version of the Latin alphabet, augmented to 43–45 letters, to write English in a (more or less) phonetically consistent way so as to make it easier for young children to learn to read and write. After initially learning to read and write using the Initial Teaching Alphabet, children then make the transition to normal English orthography.[12]

   In fact, this argument may be carried a step further, because in fact literacy rates in Sri Lanka are not merely **as good as** would be expected given the country's GDP per capita—they are **better**, 27 places better. It may be that from the point of view of literacy there is actually an unintentional side benefit to using a 'frozen' H, if the introduction of this H is delayed until fourth or fifth grade: it makes it psychologically easier to begin schooling with a maximally simple writing system based directly on the spoken language of the children who are learning to read. On the other hand, in languages which are clearly understood **not** to be diglossic, such as English or French, it is psychologically more difficult to divide the language in two in such a clear way: primary school children are immediately taught in essentially the same language as adults, which is not necessarily so easy for them. But in a diglossic language like Sinhala, it is relatively conceptually simple to expand the functional domain of L to include early literacy.[13]

---

[12] After some initial successes in the early 1960s, the Initial Teaching Alphabet was abandoned for a number of reasons. It was not sufficiently supported by either parents (who did not make the necessary effort to learn the system in order to help their children to read) or publishers (who did not publish many books using the new alphabet). The alphabet was specifically designed for children speaking Received Pronunciation, who only constitute a tiny fraction of the children in the school system, and did not take dialectal distinctions into consideration. And the transition to traditional English orthography was done much more quickly than would have been best, even in the first grade.

[13] Alexis Manaster-Ramer (personal communication) told me an interesting anecdote supporting this conclusion. Like many linguists, he had studied a number of languages from teach-yourself books and then attempted to put what he had learned into practice to talk to speakers of these languages, and also like many linguists he was frustrated to discover that speakers of e.g. French, German, etc., do not speak as the books taught—that is, the people writing the books do not feel that they should literally teach a completely colloquial version of the language. But he was quite surprised after having studied Sinhala from a book and speaking with Sinhalese that the people really did speak as the book had described—that is, the linguist writing the book really had taught the colloquial language—and it seems reasonable to attribute this to the fact that Sinhalese clearly distinguish between H and L versions of their language so that they can conceptualize actually teaching the L.

This can also be related to what seem to be remarkable data regarding literacy in ideographic languages (Chinese, Japanese, and Korean). Any adult who has studied these languages can testify that they seem to be enormously difficult to learn to read and write, because the ideographs are not obviously phonologically motivated and because so many ideographs need to be learned. Yet Japan, South Korea, and North Korea all have essentially universal literacy while the literacy rate in China (93.3 %) is **higher** than would be expected given the country's capita income (ranking 83rd in the world as opposed to a ranking of 99th in GDP per capita). How is this possible? The answer, I would argue, lies in the way that literacy is taught in these three languages, which is structured so as to make the language which primary school children learn to read first as close and as transparently related to their spoken language as possible. At the beginning of children's literacy education, all of these languages focus on consistent and simple systems of sound-based writing (the Japanese system is predominantly phonetic, the Chinese one phonemic, and the Korean one morphophonemic). Japanese uses a syllabary consisting of only 48 characters, and Korean and Chinese use basically alphabetic systems; the Korean one is purely indigenous, the only complication being that letters are organized into syllables in a conventionalized way, while the Chinese one, known as pinyin, uses a Latin alphabet designed specifically to represent Chinese sounds consistently. In all of the languages, children are introduced to ideographs gradually, as their cognitive abilities allow, and learning the ideographs is aided in the early stages by writing in small sound-based notation next to them to tell or remind the children how to pronounce them. Computers are enormously helpful in this, because even if a student only knows a particular ideograph passively, he can enter the pronunciation into the computer phonetically, and the computer will be able to turn this into the correct ideograph (the more sophisticated programs will consider the context in doing this) or at the very least give the child a few options from which to choose the correct one. This is particularly important because it is clear that children can learn to **recognize** ideographs much faster than they can learn to **write** them. Using computers in this way makes it possible for children to write everything they can say from a very early age, rather than being restricted to writing words for which they actively know the ideographs, which is very limiting and makes writing boring (Zhang and Liu n.d.).

These examples show that even in languages in which the established standard language is in one way or another not a consistent phonological representation of people's spoken language, **it is possible to get impressive literacy results from the creative use of phonologically consistent writing systems to teach basic literacy to children for the first few years of primary school**. In the case of Arabic, this would necessarily entail using different writing systems in different areas for the first few years of primary school, because of the differences between the different spoken dialects, but after a few years they could all be taught the same standard language.

In fact, as it happens, young Arabic speakers have in the last 10–15 years be-
gun to develop writing systems of this type, with the advent of electronic writ-
ing in media such as SMS and email messages, Messenger, forums, and Facebook
(see e.g. Warschauer et al. 2002; Wheeler 2003; Palfreyman and al-Khalil 2003;
Garra 2007). A typical example of this type of writing—in this case by an Arabic-
speaking citizen of Israel writing in Latin letters—is given in (1) (Garra 2007, p. 89,
90):[14]

> (1) Kolhen be2refo, bs Haifa elle 3anjad btestahal la2ano btjanen ow jamalha tabe3e.. ama
> elba2aya kolhen 3amaleyat tajmeel.. matalan dina hayek ma heye bte2ref shu 7elo feha
> ya3ne?? wala elissa mahe tomha a3waj ow mesh 7elwe shelleama zoo2 3aleko ya nas..lesh
> najwa karam 7elwe?? araaaaaaaaf!! wala amal 7ejazy mhye zai el amwat manzarha belzat
> bel look eljded!!! welko ya nas shu sayebko?? hadol!! 7elwat???? shelle la2
> They [celebrities who participated in a beauty contest]'re disgusting, Haifa [a famous Arab
> singer] deserves [to win the beauty contest] because she's gorgeous and her beauty is natu-
> ral. But all the others [singers] have had cosmetic surgery. Like Dina Hayek, what's beauti-
> ful about her? She's ugly! And Elissa, her mouth is twisted! She isn't beautiful at all! What
> kind of taste have you people got? And Najwa Karam is beautiful? Get real! Ugh! And
> Amal Hijazi!! She looks like a corpse, especially her new look. What's wrong with you
> people? These women are beautiful? Absolutely not!

Young Arabic speakers all around the Arab world are writing like this in impro-
vised writing systems, based upon their spoken dialects, using the Arabic, Latin,
or even (among Israeli Arabs) Hebrew alphabets. This is part of a general world-
wide development—young people naturally write in such contexts in their collo-
quial language, and when their colloquial language happens to be radically different
from the established written language, they write it in an improvised orthography. I
first became aware of this phenomenon in the late 1990s when a 40-year-old Swiss
woman taking one of my classes in Israel reported to me that her son back in Swit-
zerland wrote email messages to her in Swiss German dialect, which the mother of
course also spoke but which is radically different from the standard German which
has been traditionally written. I have since found that the same sort of thing is hap-
pening in diglossic languages like Persian, Bengali, and Sinhala, as well as local
dialects which differ radically from the national standard, like Sicilian in Italy and
Galician in Spain (Myhill 2009).

Although this new electronic writing was first used in a purely improvised way,
we are seeing, in the case of Arabic, the gradual development of regional norms of
writing it, based upon a combination of the local Arabic dialect and writing conven-
tions which users are agreeing upon (without, at this stage, any conscious plan-
ning). I will give here a few examples of this. For Israeli Arabs, although different
speakers can pronounce the uvular stop/q/ which the Arabic letter qaaf ق represents
as either [ʔ], [q], [k], or [g] in more personal communications they can write this
phoneme, for which no obvious Latin letter exists, as <2>, <q>, <k>, or <g>. It has
become understood in Israel in recent years that the normal 'public' written form
is <2>—but in North Africa, where the glottal stop is not used, the normal written
form for quaff is <9>, while in the Gulf area, where the normal pronunciation of qa:f

---

[14] This example was taken from the Panet forum.

is [g], this is written as <g> or <8>. On the other hand, in the Gulf area <9> is normally used to write ṣa:d ص, the emphatic voiceless alveolar fricative, while in the Levant and North Africa this phoneme is normally written as <s>, like the parallel non-emphatic. The post-alveolar voiceless fricative[š] is written <sh> by Israeli Arabs, based upon the English spelling, while Lebanese, Algerians, and Moroccans write it with <ch>, based upon French, but recently, Jordanians have found a third and monographic way to write this sound, as <$>. Similar variations are appearing in writing using the Arabic script. For example, in the Gulf, where the voiced alveo-palatal fricative/affricate ji:m is pronounced as [j], this is written with the letter ya:ʔ ي instead of ji:m ج (see discussion in Garra 2007).

These local ways of writing have not yet become fixed, but there is an unmistakable trend in this direction, particularly among users below the age of 23–24 (as I write this in 2011). Further, these conventions are almost universally known among young people who are affluent enough to have a cellular phone. Assuming that linguists in each area developed a conventionalized way to write these ʕa:mmiyyas which could be used in school, which would entail a minimal amount of work, they would be ideal as the basis for teaching early literacy to primary school children, because they are entirely phonetic and very easy to learn to read and write. In fact, this is exactly why young people have independently invented and begun to use them recently.

The next generation of Arabic teachers will come to their job already having informally learned how to write this way and will be completely used to it. Such systems for writing ʕa:mmiyya could be tailored to suit various needs. For example, in cases in which students have access to computers, computer programs could be designed which would enable the students to enter what they want to write in ʕa:mmiyya and then the computer would translate it into Fusḥa. This would make it possible for beginning students to focus on learning to **read** Fusḥa rather than learning to **write** it, which is much more difficult and frustrating (this is parallel to the way in which Chinese education policy makes use of Latin letters in teaching literacy in Chinese). In cases in which there is emphasis on learning English or French in early grades, a Latin writing system could be used which would introduce students to the Latin letters from an early age. Such steps would constitute a creative and productive approach to the serious literacy problems which are universal today in Arabic-speaking countries.

9.2   Conclusion

There is clearly an enormous literacy problem in Arabic-speaking countries. Some part of it is likely to be to due to specific features of the Arabic script (see e.g. Azzam 1984; Bentin and Ibrahim 1996; Frost et al. 1987; Ibrahim et al. 2002, 2007; and Roman and Pavard 1987), but there is every reason to believe that a very large part is due to the radical difference between the spoken Arabic dialects and the official written language which is taught in school and used for all academic materials

(Saiegh-Haddad 2003, 2004, 2007, 2012). An enormous mass of evidence from languages around the world supports the idea that children learn to read most efficiently when the language of their primary schooling is as close as possible to their native dialect—whether or not they switch afterwards to being educated in another written language. It is clear that traditional written Arabic is so different from the spoken dialects that from the point of view of learning it is effectively a foreign language. This means that we would expect that programs which attempt to teach literacy in fusḥa from first grade will have very poor results, and this is in fact what we find.

Arabic speakers are similar to many linguistic groups around the world in that they are confronted with the challenge of catching up with developed countries in terms of literacy rates, and they would do well to observe which tactics have been successful and which tactics have been unsuccessful for other groups facing similar situations. Among those groups which have been successful, the ex-Soviet groups and the Slavic groups devised writing systems based directly and closely on the local spoken language, in the process making up a large number of new written languages. The Chinese, Japanese, and Koreans have continued to use their traditionally ideographic writing systems but have also made extensive use of sound-based writing systems corresponding to the spoken language in primary education, and in the case of the Chinese they have even used the Latin alphabet to do this. The Sinhalese have maintained their traditional H language but have introduced a written version of their L language for the first four grades of primary school. These programs have varied in a number of regards but they have had one thing in common: they have all instituted programs in which, in the first few years of schooling, children are taught to read and write in a phonologically transparent writing system which is based directly upon the spoken language which they have already learned.

On the other hand, there are two approaches which have been extremely **un**successful in terms of promoting literacy. One is the retention of foreign ex-colonial languages, in particular English and French, as the language of education even in countries where people do not speak English or French as their native language. The second is the use of a 'frozen' H language as the language of education, as in Arabic, Persian, and a number of diglossic languages of the Indian subcontinent (but not Sinhala).

The obvious conclusion to draw from this is that Arabic speakers would be well-advised to switch from the latter type of strategy to the former. The cases of Sinhalese, Chinese, Japanese, and Korean have shown that it is possible to do this and to radically increase literacy rates while at the same time maintaining traditional and highly-valued standard written languages, if a certain amount of creative ingenuity is employed. The new writing systems which young Arabic speakers have recently been developing for use in electronic media would seem to be ideal—and their construction at the present time highly fortuitous—for this purpose, but use of these systems in primary education is only one possible way of making Arabic literacy programs more efficient. What is really important is the general principle: literacy is best taught by **beginning** with a sound-based writing system which is as close as possible to the spoken language which the child already knows

References

Abu-Rabia, S. (2000). Effects of exposure to literary Arabic on reading comprehension in a diglossic situation. *Reading and Writing, 13,* 147–157.

Alidou, O. (1996). Francophonie, World Bank, and the collapse of the Francophone Africa educational system. *CAFA Newsletter*, 11.

Azrael, J. R. (Ed.). (1978). *Soviet nationality policies and practices*. New York: Praeger Publishers.

Azzam, R. (1984). Orthography and reading of the Arabic language. In J. Aaron & R. M. Joshi (Eds.), *Reading and writing disorders in different orthographic systems*. Kluwer Academic.

Bentin, S., & Ibrahim, R. (1996). New evidence for phonological processing during visual word recognition: The case of Arabic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 309–323.

Berger, M. (1990). Diglossia within a general theoretical perspective: Charles Ferguson's concept 30 years later. *Multilingua, 9,* 285–295.

Bokamba, E. G. (1984). French colonial language policy and its legacy. *Studies in Linguistics*, *14*.

Bokamba, E. G., & Tlou, J. (1997). The consequences of the language policies of African states vis- à-vis education. *Proceedings of the VII Conference on African Linguistics*. Columbia: Hornbeam.

Browning, R. (1982). Greek diglossia yesterday and today. *International Journal of the Sociology of Language, 35,* 49–68.

Calvet, L.-J. (1974). *Linguistique et Colonialisme: Petit Traité de Glottophagie*. Paris: Payot.

Daltas, P. (1993). The concept of diglossia from Ferguson to Fishman to Fasold. In I. Philippaki-Warburton, K. Nicolaidis, & M. Sifianou (Eds.), *Themes in Greek linguistics: Papers from the first international conference on Greek linguistics*. Amsterdam: Benjamins.

Del Castillo, D. (2001). The Arabic publishing scene is a desert, critics say. *Chronicle of Higher Education*, *47*(48), A55–A57.

Doets, C. (1994). Assessment of adult literacy levels: The Dutch case. In Verhoeven. (Ed.), *Functional literacy* (pp. 321–332). Amsterdam: John Benjamins.

Downing, J., & Lathan, W. (1967). *Evaluating the initial teaching alphabet: A study of the influence of English orthography in learning to read and write*. London: Cassell.

Dumont, P., & Maurer, B. (1995). *Sociolinguistique du Français en Afrique Francophone*. Vanves: EDICEF.

Dutcher, N., & Tucker, G. R. (1997). *The use of first and second languages in education: A review of educational experience*. Washington D.C.: World Bank, Country Department III.

Eckert, P. (1980). Diglossia: Separate and unequal. *Linguistics, 18,* 1053–1064.

Fellman, J. (1975). On 'diglossia.' *Language Sciences, 34,* 38–39.

Ferguson, C. (1959). Diglossia. *Word, 15,* 325–340.

Ferguson, C. (1991). Diglossia revisited. *Southwest Journal of Linguistics, 10,* 214–234.

Fishman, J. A., Conrad, A. W., & Rubal-Lopez, A. (Eds.). (1996). *Post-imperial English: Status change in former British and American colonies, 1940–1990*. Berlin: Mouton de Gruyter.

Frangoudaki, A. (1992). Diglossia and the present language situation in Greece: A sociological approach to the interpretation of diglossia and some hypotheses on today's linguistic reality. *Language in Society, 21,* 365–381.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 104–114.

Garra, E. (2007). *From a dialect into a language: The cases of English and Arabic*. MA thesis, University of Haifa English Department.

Gill, H. (1999). Language choice, language policy and the tradition–modernity debate in culturally mixed postcolonial communities: France and the francophone Maghreb as a Case study. In Y. Suleiman (Ed.), *Language and society in the middle east and north Africa*. Richmond: Curzon Press.

Gray, W. S. (1956). *The teaching of reading and writing*. Chicago: Scott Foresman.

Gudschinsky, S. C. (1977). Techniques for functional literacy in indigenous languages and the national language. In T. P. Gorman (Ed.), *Language and literacy: Current issues and research*. Teheran: International Methods.

Haeri, N. (2009). The elephant in the room: Language and literacy in the Arab world. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy*. Cambridge: Cambridge University Press.

Hudson, A. (2002). Outline of a theory of diglossia. *International Journal of the Sociology of Language, 157,* 1–48.

Ibrahim, R., & Aharon-Peretz, J. (2005). Is literary Arabic a second language for native Arab speakers? Evidence from a semantic priming study. *The Journal of Psycholinguistic Research, 34*(1), 51–70.

Ibrahim R., Eviatar, Z., & Aharon-Peretz, J. (2002). The characteristics of Arabic orthography slow its processing. *Neuropsychology, 16,* 322–326.

Ibrahim, R., Eviatar, Z., & Aharon-Peretz, J. (2007). Metalinguistic awareness and reading performance. *Journal of Psycholinguistic Research, 36,* 297–317.

Khamis-Dakwar, R. (2005). Children's attitudes towards the diglossia situation in Arabic and its impact on learning. *Language, Communities, and Education, 1,* 75–86.

Khamis-Dakwar, R. (2007). *The development of diglossic morphosyntax in Palestinian Arabic-speaking children*. Unpublished dissertation, Columbia University, New York.

Levine, K. (1994). Functional literacy in a changing world. In L. Verhoeven (Ed.), *Functional literacy: Theoretical issues and educational implications* (pp. 113–131). Amsterdam: John Benjamins.

Lewis, E. G. (1972). *Multilingualism in the Soviet Union: Aspects of language policy and its implementation*. The Hague: Mouton.

Maamouri, M. (1998). *Language education and human development: Arabic diglossia and its impact on the quality of education in the Arab region*. Washington, D.C.: The World Bank.

Mehrotra, S. (1998). *Education for all: Policy lessons from high-achieving countries*. New York: UNICEF Staff Working Papers.

Myhill, J. (2009). Towards an understanding of the relationship between diglossia and literacy. Hayozma lemexkar yisumi bexinux, on the website of the Israeli Academy. http://www.academy.ac.il/data/projects/34/diglossia_myhill_Eng_May2009.pdf.

Okedara, J. T., & Okedara. C. A. (1992). Mother tongue literacy in Nigeria. *Annals AAPSS, 520,* 91–102.

Ornstein, J. (1968). Soviet language policy: Continuity and change. In E. Goldhagen (Ed.), *Ethnic minorities in the Soviet Union*. New York: Frederick A. Praeger.

Palfreyman, D., & el-Khalil, M. (2003). 'A funky language for teenzz to use': Representing Gulf Arabic in instant messaging. *Journal of Computer- Mediated Communication*, *9*.

Phillipson, R. (1992). *Linguistic imperialism*. Oxford: Oxford University Press.

Pool, J. (1978). Soviet language planning: Goals, results, options. In J. P. Azrael (Ed.), *Soviet nationality policies and practices*. New York: Praeger Publishers.

Powell, R. (2002). Language planning and the British empire: Comparing Pakistan, Malaysia and Kenya. *Current Issues in Language Planning, 3,* 205–278.

Roman, G., & Pavard, B. (1987). A comparative study: How we read Arabic and French. In J. K. O'Regan & A. Levy-Schoen (Eds.), *Eye movement: From physiology to cognition*. Amsterdam: North Holland Elsevier.

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics, 24,* 431–451.

Saiegh-Haddad, E. (2004). The impact of phonemic and lexical distance on the phonological analysis of words and pseudo-words in a diglossic context. *Applied Psycholinguistics, 25,* 495–512.

Saiegh-Haddad, E. (2007). Linguistic constraints on children's ability to isolate phonemes in Arabic. *Applied Psycholinguistics, 28,* 605–625.

Saiegh-Haddad, E. (2012). Phonological processing in diglossic Arabic: The role of linguistic distance. In E. Broselow & H. Ouli (Eds.), *Perspectives on Arabic Linguistics XXII* (pp. 269–280). John Benjamins Publishers.

Saiegh-Haddad, E., Levin, I., Hende, N., & Ziv, M. (2011). The linguistic affiliation constraint and phoneme recognition in diglossic Arabic. *Journal of Child Language, 38,* 297–315.

Salhi, K. (2002). Critical imperatives of the French language in the Francophone world: Colonial legacy–postcolonial policy. *Current Issues in Language Planning, 3,* 317–345.

Schiffman, H. (1997). Diglossia as a sociolinguistic situation. In F. Coulmas (Ed.), *The handbook of sociolinguistics*. Oxford: Blackwell.

Scotton, C. (1986). Diglossia and code switching. In J. Fishman et al. (Eds.), *The Fergusonian impact: In honor of Charles A. Ferguson on the occasion of his 65th birthday*, vol. 2: *Sociolinguistics and the sociology of language*. Berlin: Mouton de Gruyter.

Simon, G. (1991). *Nationalism and policy toward the nationalities in the Soviet union: From totalitarian dictatorship to post-Stalinist society* (trans: K. Forster & O. Forster). Boulder: Westview Press.

UNESCO. (1953). *The use of vernacular languages in education*. Paris: UNESCO.

Verhoeven, L. (1994). Modeling and promoting functional literacy. In L. Verhoeven (Ed.), *Functional literacy: Theoretical issues and educational implications*. Amsterdam: John Benjamins.

Verhoeven, L. (1997). Acquisition of literacy by immigrant children. In C. Pontecorvo (Ed.), *Writing development: An interdisciplinary view*. Amsterdam: John Benjamins.

Wagner, D. A. (1993). *Literacy, culture, and development: Becoming literate in Morocco*. Cambridge: Cambridge University Press.

Wagner, D. A., Spratt J. E., & Ezzaki, A. (1989). Does learning to read in a second language always put a child at a disadvantage? Some counter-evidence from Morocco. *Applied Psycholinguistics, 10,* 31–48.

Wexler, P. (1971). Diglossia, language standardization, and purism: Parameters for a typology of literary languages. *Lingua, 27,* 330–354.

Wheeler, D. L. (2003). The internet and youth subculture in Kuwait. *Journal of Computer-Mediated Communication*, *8*.

Warschauer, M., El Said, G., & Zohry, A. (2002). Language choice online: Globalization and identity in Egypt. *Journal of Computer-Mediated Communication*, *7*.

Zhang, D., & Liu, Y. (n.d.). Pinyin input experiments in early Chinese literacy instruction in China: Implications for Chinese curricular and pedagogic reform in Singapore. http://conference.nie.edu.sg/paper/Converted%20Pdf/ab00256.pdf.